This product is part of the RAND Corporation monograph series. RAND monographs present major research findings that address the challenges facing the public and private sectors. All RAND monographs undergo rigorous peer review to ensure high standards for research quality and objectivity.

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

Gail L. Zellman, Michal Perlman, Vi-Nhuan Le, Claude Messan Setodji

Prepared for Qualistar Early Learning

RAND EDUCATION

The research described in this report was prepared for Qualistar Early Learning and was conducted by RAND Education, a unit of the RAND Corporation.

Library of Congress Cataloging-in-Publication Data

Assessing the validity of the Qualistar early learning quality rating improvement system as a tool for improving child-care quality / Gail L. Zellman ... [et al.].
p. cm.
Includes bibliographical references.
ISBN 978-0-8330-4495-2 (pbk. : alk. paper)
1. Child care—United States—Evaluation. 2. Child care services—United States—Evaluation. 3. Early childhood education—United States—Evaluation. I. Zellman, Gail.

HQ778.63.A77 2008
362.71'20684—dc22

2008022083

Published 2008 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

The generally low quality of child care in the United States has led to calls to improve quality. One approach to quality improvement that has been embraced widely of late involves the development and implementation of quality rating systems (QRSs), multicomponent assessments designed to make child-care quality transparent and easily understood. Participating providers are assessed on each of the system components and receive a summary rating that they are encouraged to display. In theory, these simple, readily understood ratings (often 0–5 stars, or a rating of 1–4) will enable parents, funders, and other stakeholders to make more informed choices about which providers to use and support, and will encourage providers to improve. Quality rating and *improvement* systems (QRISs) also include feedback, technical assistance, and incentives to both motivate and support quality improvement.

A key limitation on QRSs and QRISs is the lack of data about these systems—how well they measure what they purport to measure, 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.

Qualistar Early Learning, a Colorado-based nonprofit organization, was one of the first to create a QRIS. Qualistar approached RAND in 2000, asking for help in evaluating the validity of the Qualistar QRIS. RAND assessed the five Qualistar QRIS components separately, then examined how they related to each other; compared Qualistar QRIS measures to other, established measures of quality; and examined whether quality improvements as measured by the Qualistar QRIS components were associated with better child outcomes.

This report describes the results of our work, conducted from 2000 to 2007. It should be of interest to early childhood educators and policymakers concerned with improving child-care quality and to researchers working to develop better measures of care quality. Elements of this work may be found in Perlman, Zellman, and Le (2004), Le et al. (2006), and Zellman and Perlman (2006).

This study was carried out by RAND Education, a unit of the RAND Corporation, and was funded by Qualistar Early Learning. The study reflects RAND Education's mission to bring accurate data and careful, objective analysis to the national discussion on early care and education. Any opinions, findings, and conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of Qualistar Early Learning.

Contents

Figures

Tables

Summary

Introduction

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. One increasingly popular approach involves developing and implementing what are known as quality rating systems, or QRSs.

QRSs use multicomponent assessments to produce a summary quality rating designed to make child-care quality transparent and easily understood. In reality, most QRSs are really QRISs—quality rating and *improvement* systems. QRISs have been promoted because it became apparent early on that motivation alone was not enough to enable providers to improve quality. Many child-care providers lack the ability to determine how to improve; the summary ratings that are the product of the QRS assessment process provide little guidance. QRISs provide more-detailed assessments, hands-on technical assistance, and quality-improvement resources to participating providers to improve the level of quality they offer.

Despite their growing popularity, there is little information available about how well QRISs work. A logic model presented in this report posits a clear path to improved provider quality and better child outcomes, but it is largely untested. We do not know how well QRISs 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, or whether children benefit from the improved care they are receiving as their provider receives quality-improvement support.

We do know that QRISs need attention. A number of widely used QRIS component measures were designed for use in research studies and quality-improvement efforts. These are what can be called "low-stakes" contexts, because the findings have few implications for the programs being assessed. Of late, the stakes are getting higher, with people talking about and, in some cases, actually rewarding higher-quality child-care programs with higher per-child subsidies and other incentives. Quality measures created for low-stakes contexts are not necessarily considered valid in "high-stakes" contexts; they need to be validated in the high-stakes contexts in which they will be used.

Recognizing these challenges and gaps, Qualistar Early Learning—a Colorado-based nonprofit and one of the first child-care organizations to create a QRIS, which was first implemented in 1999—asked RAND to assess how effectively its QRIS was working. Its rating system, which we call the Q-QRIS to distinguish it from others' systems, includes five components generally agreed to contribute to high-quality care: classroom environment, child-staff ratios, staff and director training and education, parent involvement, and accreditation. Up

to 10 points are awarded for every component except accreditation, which is worth 2 points. Thus, any given program can receive a maximum of 42 points. Those points are converted to a star rating of 0–4 stars, depending on the number of points received.

To assess the Q-QRIS, we examined 65 child-care centers and 38 family child-care providers (which provide child care in family homes) using the Q-QRIS as well as two other measures of quality: the Caregiver Interaction Scale (CIS) and the Pre-Kindergarten (Pre-K) Snapshot. One preschool-age classroom per child-care center was examined in depth, although the Q-QRIS components were measured in all classrooms. We assessed the social, emotional, and cognitive functioning of participating children based on teacher surveys and direct examination and also collected extensive family background information from parents. All told, we collected data on over 1,300 children in the first wave of data collection. The same instruments were administered over two additional waves of data approximately 12 months apart.

It is important to note that while 75 percent of the original 65 centers remained through all three study waves, those that left the study after Wave 1 were lower in quality than centers that remained. Child attrition within the centers themselves was an even more serious problem; only 7 percent of the original sample remained in the study through Wave 3. Also, several component measures changed over time as well. These circumstances have affected the study findings, as we discuss below.

Findings

Our assessment addressed seven questions, which are discussed below.

1. What Are the Characteristics of the Q-QRIS Components as Measures?

As noted previously, the Q-QRIS consists of five components; one of the goals of the evaluation was to understand the properties of these measures and determine how well they assess quality. Our early analyses identified *significant measurement issues with several of the Q-QRIS components.* As a result, we devoted a great deal of effort during our evaluation to improving these measures.

In particular, we examined the measurement of *child-staff ratios* by examining a unique measurement approach and comparing it with others. Specifically, we used sign-in/sign-out data (i.e., where staff sign children in and out of rooms) from 77 centers. Doing so told us that it is important to collect ratios over a period longer than the typical collection timeframe (two hours in the morning) and to collect ratios from more than one classroom serving a particular age group, since ratios differ across such classrooms. Qualistar replaced self-reports of ratios with observations as the data-collection method, since the former are untrustworthy in high-stakes settings. Our analyses informed the frequency of these observations.

Our analyses showed that *parent involvement* was poorly measured at the outset of our study, since programs that varied substantially in quality were all rated very highly by the parents whose children were cared for in them. Several subsequent parent involvement measures we introduced also produced very limited response variability despite a wide range in program quality. Ultimately, the Family Partnership (FP) measure was developed, which is based on the notion that productive parent child-care involvement efforts should focus on helping parents to develop and maintain a good relationship with their child. Qualistar adopted this measure, which both parents and providers complete. We found some response variation across

programs on the FP measure; it also related to some other quality measures. However, more research on this measure is needed.

We examined the psychometric properties of the Early Childhood Environment Rating Scale–Revised (ECERS-R). We found that the 43 items that compose this scale were highly correlated, suggesting that it may be possible to reduce the number of items that are administered. We also found that ECERS-R ratings across classrooms within a center were highly correlated, suggesting that it may be possible to assess only some of the classrooms and still capture the quality of all rooms with reasonable accuracy. Finally, we found that the ECERS-R captures one global aspect of quality rather than the seven scales outlined by the measure developers. This indicates that only the total score should be used, as is the practice with the Q-QRIS.

Teacher training and education measures still need a good deal of attention; we will be focusing on the assessment of movement of staff over the course of the day in future work because we believe it is fundamental to assessing and understanding the importance of staff training and education. If teachers (or children) do not remain in a given classroom for very long, it is not clear how to assess the effect of teacher background on classroom processes or children's outcomes.

Finally, we found limited relationships between *accreditation* status and other measures of quality. Given this finding, Qualistar should consider whether the cost and effort required for providers to earn national accreditation is justified.

2. How Closely Related Are the Five Q-QRIS Component Measures?

Beyond trying to understand how well the Q-QRIS component measures work as measures, we also wanted to know how well they correlate with each other. Since all components assess child-care quality, there should be some relationships. However, since each component purportedly measures a different aspect of quality, they should not relate too closely.

When we looked across all three waves of the data, we found that *the component measures correlated moderately well*. In particular, lower child-staff ratios and better education and experience among head teachers and directors are associated with higher-quality classroom environments. We also found that accreditation is associated with higher scores on the ECERS-R, which the Q-QRIS relies on to assess the classroom environment in centers. Finally, FP parent and provider points are positively associated with head teacher and director education and negatively associated with child-staff ratios, as expected.

3. Do Providers That Receive High Scores on the Star Ratings and the Individual Q-QRIS Components Also Receive High Scores on Process-Quality Measures?

Child-care quality is generally viewed as encompassing both *structural* characteristics, such as ratios and staff training and education, and *process* characteristics, which 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 than structural quality because it influences children more directly. Structural characteristics, such as those measured in the Q-QRIS, are viewed as driving the quality of the processes that take place in a given setting.

Because of this, we wanted to examine the relationships between the star ratings that are the ultimate output from the Q-QRIS, the individual Q-QRIS component measures that yield

those star ratings, and the process-quality measures. We selected two commonly used process-quality measures for this purpose: the CIS and the Pre-K Snapshot.

We found that, at Wave 1, providers that scored high on the Q-QRIS components and star ratings scored high on a few of the CIS subscales, and virtually all these relationships were in the expected direction. However, there were no significant associations between the star ratings or the Q-QRIS components and the Pre-K Snapshot. Components that focused more directly on process quality were more closely associated with the process-quality measures, while those that were most structural (e.g., child-staff ratios) demonstrated the fewest relationships. However, these findings were not replicated across Waves 2 and 3. Lack of consistency in findings across waves may be at least partly the result of that fact that attrition from the study sample was not random across the waves; lower-quality providers were more likely to drop out over the course of the study. *Taken together, the results suggest that the star ratings and the Q-QRIS components are generally unrelated to measures of process quality.*

4. Is There a Relationship Between the Q-QRIS Components and Concurrent Child Outcomes? Is Provider Quality Related to Future Child Outcomes?

According to the logic model underlying QRISs, an improved child-care environment, characterized by more responsive caregiving and enriched content, will 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.

Given this, we examined the relationship between the star ratings, Q-QRIS components, and child outcomes. *We found few relationships between individual Q-QRIS components and child outcomes and virtually none between star ratings and child outcomes. As with the process-quality correlations, the results were not replicated across waves.*

We found a very similar pattern of results for the cross-sectional analyses that used the two process-quality measures to predict child outcomes. The four subscales for each of these two measures did not predict any of the child outcomes.

As noted above, we also looked at family child-care settings as part of the study. The family child-care data, based on small numbers, found a few relationships between the star ratings and the individual Q-QRIS components and both the process-quality measures and child outcomes. However, here, too, the effects were inconsistent across waves.

5. How Should the Components Be Combined into a Q-QRIS to Account for the Relative Contributions of the Components to Child Outcomes?

Combining quality components that measure quite different aspects of quality is a key Q-QRIS innovation, and we hoped to be able to provide empirical guidance on how to do it in a way that best assessed quality. *However, because we did not find large or consistent relationships between the Q-QRIS components and child outcomes, we were unable to address issues of combining or weighting the Q-QRIS components.*

6. Are There Subgroups of Children for Whom the Links Between Measures of Child-Care Quality and Child Outcomes Are Stronger?

Although we did not find any strong and consistent links between the star ratings or the Q-QRIS components and improved child outcomes in the general population, this does not mean that such links will not show up among subgroups within that population. The strongest effect sizes in the literature on the impacts of quality child care and improved child outcomes

are reported for studies in which disadvantaged children are randomly assigned to programs that provide high doses of standardized, high-quality care and extensive support to parents in a very prescribed way.

To determine whether certain subgroups of children were affected differently from the general population, we conducted a series of parallel analyses with subgroups of children who came from low-income homes or who had experienced high doses of child-care exposure. We found that *the pattern of results for these children did not differ from that found for the general population*.

7. Did Center Quality Change Over Time? Did Family Child-Care Quality Improve Over Time? If So, Did the Q-QRIS Components Also Improve Over Time?

The final question we were interested in answering had to do with whether child-care quality improved over time. We found that *provider quality did improve.* One indicator of improvement was the increase in the percentage of accredited centers over the course of the study. Family child-care provider quality also improved slightly.

However, it is not possible to unequivocally attribute these changes to the Q-QRIS. It is possible that improvements were simply a reaction to being assessed or were part of regular practice in a group of providers that self-selected into a quality-improvement study. Intervention participant self-selection, the lack of a comparison group, and limited data on the implementation of the intervention made testing the effect of the intervention impossible.

Summary of Findings

The seven questions above provide Q-QRIS validity evidence. Taken together, the findings provide mixed support for the Q-QRIS and its components as measures of provider quality. The Q-QRIS and the component measures correlate moderately with each other and show some relationships with one of the two process measures chosen as criterion measures. Although the underlying logic model suggests that the Q-QRIS and its components should predict child outcomes, we found little evidence to support these relationships.

Definitive conclusions about the validity of the Q-QRIS and its components cannot be drawn because of study design and implementation limitations, including criterion measures collected from a single classroom in each center; ECERS-R data primarily collected in settings where stakes were not attached to scores; a new measure of parent involvement that showed promise because it produced variation in responses across programs of varying quality, but has yet to be validated; lack of a randomized design; nonrandom provider attrition; and very high child attrition in our sample. These limitations also make it difficult to generalize our findings to the functioning of similarly constructed QRISs in other settings.

Implications

Our findings raise a number of important questions; key among them are (1) what we learned about implementing QRISs at scale from our work with the Q-QRIS and (2) what to make of the lack of association we find between the Q-QRIS and child outcomes.

Implementing QRISs at Scale

As accountability increasingly becomes a driving concept in American education, quality rating systems are proliferating in the child-care arena. But virtually no one is focusing serious effort on how to build good systems. Little attention is being paid to determining which components are most important, how best to measure them, or how to weight and combine the component measures to produce the summary ratings that characterize these systems.

Qualistar Early Learning is to be applauded for understanding that good quality rating systems must be evidence-based; built on careful empirical analyses of component measures; revised as needed to improve those measures; and weighted and combined based on empirical data to produce meaningful, defensible ratings. This study reminds us that building QRISs is a challenging task. As more states adopt them, it is also becoming an increasingly important task.

More specifically, this study's findings clearly indicate that much work needs to be done before we can confidently design and implement quality rating systems at scale. As a starting point, a research base must be established that provides data on how to best measure individual components, which components matter most, and how component scores should be combined and weighted to produce the summary ratings that are the key output of these systems. This study produced valuable information about the measurement of some key components; much more work is needed.

This focus on the Q-QRIS components points to an important lesson about quality rating systems: Building a QRIS takes time and probably, for efficiency's sake, should be done incrementally. Each construct to be measured must be clearly articulated, designed, tested, and validated in the context in which it will be used. Once the components are well measured, an iterative, evidence-based validation process on the QRIS as a whole can begin. A focus on measurement research will slow the rollout of quality rating and improvement systems, but we believe the delay will produce better systems.

These findings have led us to work with other stakeholders to develop a QRIS consortium that would devote resources to sharing data and conducting the many research studies that are required to provide an empirical basis for QRISs. Such research would make these systems more defensible, enable system developers to create more efficient measures of the key components that underlie these systems, and focus attention on attainable QRIS outcomes. Given the increasing amount of resources directed to these systems and the high stakes attached, such work is critical if we are to ensure that providers, children, and families benefit as much as possible from QRISs.

Relationships Between Q-QRIS and Child Outcomes

As noted above, our study did not find a strong and consistent link between the Q-QRIS and child outcomes. To understand how our findings compare with those of other research studies (i.e., to "contextualize" our findings), we conducted a targeted literature review of the few studies that allow direct comparison with our data by presenting analyses of associations between one or more Q-QRIS components and child outcomes; that so few studies allow for direct comparison with our data is not terribly surprising given that there are many ways to measure quality and analyze study findings. In analyzing the small set of analogous studies, we found mixed results that are consistent with what we found in this study; while some studies report significant relationships between components that are found in the Q-QRIS and child outcomes, some find no relationships at all.

It is possible that the lack of relationships between child-care quality measures and child outcomes reflects poor measurement of key components of quality. For example, to our knowledge, our work on measuring child-staff ratios provides the first empirical basis for assessing the validity of procedures to capture such ratios. Thus, estimates of ratios generated in past studies may not have captured the construct well.

In our analysis, we could not address the magnitude of the relationships that do exist. However, as discussed above, effect sizes appear to be small, even when they are statistically significant. Studies that found both significant and nonsignificant effects were most likely to have samples that included children of different income levels; a lack of relationships between quality indicators and child outcomes was most common in studies that relied on more-affluent samples. This conclusion is consistent with the notion supported by major longitudinal studies that child-care quality is most likely to influence the functioning of less privileged children.

While it makes sense and holds general appeal that improved quality will translate into improved child outcomes, the many factors that shape children over time may swamp the association, at least in the short term. The major longitudinal studies—the Carolina Abecedarian Project and the High/Scope Perry Preschool Project—find child effects many years later when contrasted with no intervention at all. But these studies provided intensive interventions to very needy children using stronger methodologies (i.e., random assignment) than our study allowed. Most child-care settings do not provide a standardized intervention, and evaluations of outcomes do not compare a no-treatment condition against a standardized one. These differences may explain the weaker pattern of findings reported here.

Despite the above caveats, it still remains that close examination of studies that are most comparable with this study suggests that the logical and appealing assumption that child-care quality is associated with improved child outcomes may not have empirical support. This finding, in turn, raises the broader question of which QRIS outcomes are the most reasonable to expect.

For example, should we expect that many of the child-functioning measures, which we know to be heavily influenced by family and child factors, will be affected by what child-care providers do? Are there better indicators of child functioning on which we can base a quality rating system? Early childhood educators, researchers, and kindergarten teachers are more interested in children's capacity to regulate their emotions, develop trusting relationships with adults, and approach learning in a motivated, efficacious way than they are in whether children acquire pre-academic skills. Should we develop and employ more of these sorts of indicators in our examinations of quality rating systems? Alternatively, it may be appropriate, particularly until we can build a stronger empirical basis for our quality measures, to stay away from longer-term child outcomes entirely, focusing instead on program outputs, such as children's engagement in developmentally appropriate tasks in a safe and supportive environment. Analogously, it may be best to focus on formative evaluations rather than summative ones until we know more about component measurement and its aggregation into summary ratings. Clearly, more research should be directed to these efforts.

Acknowledgments

This work was sponsored by Qualistar Early Learning. We are particularly grateful for the vision of Doug Price, the founder of Qualistar Early Learning, who believed that a quality rating system would help improve child-care quality for Colorado parents and children and had the understanding and courage to insist that the system be tested and refined to be as good as possible. We very much appreciate the encouragement and support of Gerrit Westervelt, now of BUILD; Kirk Williams, University of California, Riverside; and Paula Neth and Diana Schaack of Qualistar, who were always willing to talk through challenges and provide important insights into how the rating system was functioning for child-care providers and staff. Members of the Evaluation Oversight Committee were Suzanne Helburn, University of Colorado, Denver; Norman Watt, University of Denver; Susan Downs-Karkos, the Colorado Trust; Nancy Csuti, the Colorado Trust; Meera Mani, the Clayton Foundation; and Lynn Andrews, the Clayton Foundation. Committee members made themselves available whenever they were needed and provided informed and useful insights into the work. We also appreciate Carollee Howes's contributions to the measurement selection process and her willingness to make herself available for consultation at many points in the study.

We owe a huge debt of gratitude to Lynda Fernyhough of the University of Toronto for her devotion to this work. We are also grateful to Chung Pham for analytic assistance and to Christopher Dirks for his indispensable assistance on this report and throughout the project. Kevin Kapeller at OMNI Institute provided invaluable assistance with data collection and training. The report benefited from reviews provided by Linda Smith of the National Association of Child Care Resource and Referral Agencies (NACCRRA); Felipe Martinez of RAND and the University of California, Los Angeles; and Rebecca Kilburn of RAND.

Finally, we are indebted to the child-care directors and staff who allowed us to study their every move, the parents who completed our surveys and allowed us to study their children, and to the children themselves.

Abbreviations

A.A.	associate of arts degree
B.A.	bachelor of arts degree
CBI	Child Behavior Inventory
CDA	child development associate's degree
CHIP	Center for Human Investment Policy
CIS	Caregiver Interaction Scale
CLASS	Classroom Assessment Scoring System
CRPR	Child Rearing Practices Report
ECCRN	Early Child Care Research Network
ECE	Early Childhood Education
ECERS	Early Childhood Environment Rating Scale
ECERS-R	Early Childhood Environment Rating Scale–Revised
ECRP	Early Childhood Research and Practice
FBQ	Family Background Questionnaire
FDCRS	Family Day Care Rating Scale
FDR	False Discovery Rate
FP	Family Partnership
HB	House Bill
ITERS	Infant/Toddler Environment Rating Scale
Max	maximum range
Min	minimum range
NACCRRA	National Association of Child Care Resource and Referral Agencies

NAEYC	National Association for the Education of Young Children
NICHD	National Institute of Child Health and Human Development
NLS-Y	National Longitudinal Survey of Youth
NS	not statistically significant
OCS	Organizational Change Survey
ORCE	Observational Record of the Caregiving Environment
PPVT	Peabody Picture Vocabulary Test
Pre-K	pre-kindergarten
QI	quality improvement
Q-QRIS	Qualistar quality rating and improvement system
QRIS	quality rating and improvement system
QRS	quality rating system
SD	standard deviation; standard deviation unit
SDQ	Strength and Difficulties Questionnaire
T&E	training and education
WJ	Woodcock-Johnson
WJ – AP	Woodcock-Johnson Applied Problems subtest
WJ – LWI	Woodcock-Johnson Letter-Word Identification subtest
WJ – PC	Woodcock-Johnson Passage Comprehension subtest

CHAPTER ONE

Introduction

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

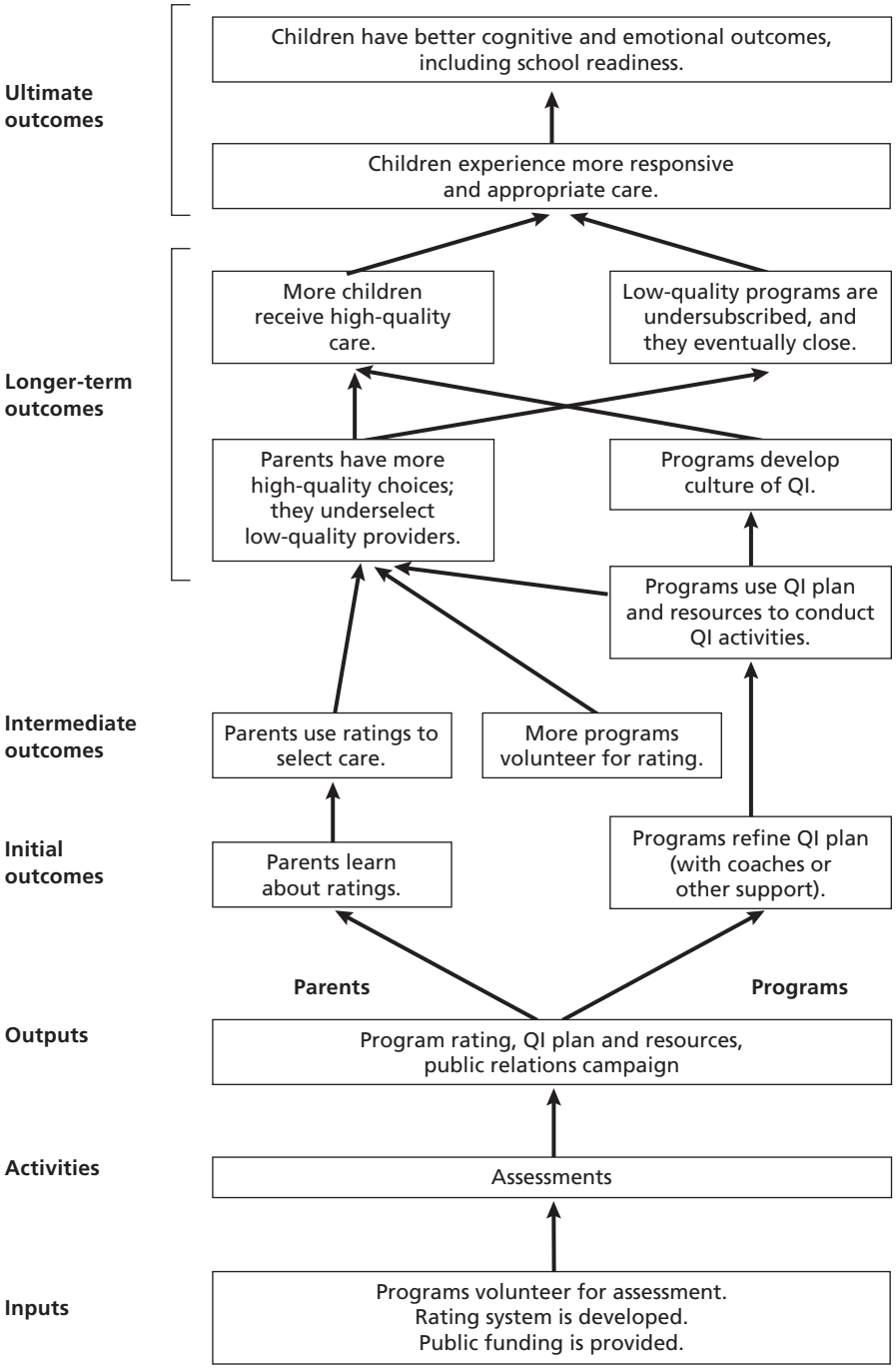
Although the challenges are daunting, there have been a number of attempts to intervene in the child-care system to promote quality improvements. One approach that has been embraced widely of late involves the development and implementation of quality rating systems (QRSs). (See Zellman and Perlman, 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 (NACCRRA, 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 Buelle 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.

CHAPTER TWO

Methods

The goal of this study was to assess the validity of the Qualistar quality rating and improvement system. Validation of the Q-QRIS involved three main tasks:

1. examination of properties of each of the system components
2. assessment of the relationship of the components to each other and to other measures of quality
3. assessment of the relationship between the system components and the summary ratings and child functioning, concurrently and across time.

To test the individual components, we drew, at times, from other sources of data, as described in Chapter Three. This chapter describes the methods used to address the primary goal of this study: testing the validity of the Q-QRIS as a whole.

Most of the providers were recruited through Colorado House Bill (HB) 1297, state legislation designed to improve school readiness and child-care quality in communities with low-performing elementary schools.[1] Providers were assessed three times, at approximately 12-month intervals. At each wave, extensive provider quality and child assessments were conducted. Providers were then given detailed feedback about their performance on each of the Q-QRIS components and some support to improve (a median of $3,500 per classroom per year). The first wave served as a measure of baseline provider quality and child functioning. The design described below allowed for both cross-sectional and longitudinal comparisons.

The Study Sample

Participating Providers

In total, 65 centers and 38 homes joined the study for the Wave 1 assessment. (See Table A.2.1 in the appendix to this chapter for a list of participating provider characteristics.)

[1] To be included in HB 1297, communities were required to submit a grant to the Colorado Department of Human Services addressing how they were going to improve children's school readiness and child-care quality. Local communities selected child-care programs to participate in the program that were within close proximity to low-performing elementary schools and served at least 50 percent low-income children. Participating programs received approximately $3,500 per classroom per year for three years. The degree of control over the HB 1297 money varied by community. Some classrooms could spend their funding at their discretion to meet the goals of their quality-improvement plan, while others were more constrained. For example, some communities were required to hire mentors to go into programs to provide technical assistance to HB 1297 classrooms. HB 1297 did not require programs to track how they spent their funding, so no information is available about how funds were allocated.

Recruitment of Child-Care Providers. Providers were recruited through a number of Colorado quality-improvement initiatives. Most centers were receiving funds through HB 1297, Colorado legislation designed to improve school readiness in communities with low-performing schools. The first group of providers were part of the original study design (see Document A.2.1 in the appendix to this chapter for a description of that design) and were recruited, in part, based on their county's willingness to provide funds for differential reimbursement to improve staff salaries.[2]

Qualistar also received discretionary monies from the Donner Family Fund, which were used to recruit centers that served more-affluent families. To find these centers, Qualistar contacted all university lab schools within a 75-mile radius of Denver that served populations that included no more than 30 percent of families who were eligible for the Colorado Child Care Assistance program. Of the six programs that were eligible, four agreed to participate in the evaluation. These programs did not receive mentoring or technical assistance funding. However, they did receive approximately $750 per classroom per year for each of the study's three years. In all cases, programs participating in the study had to serve 2.5- to 5-year old children on a full-time basis.

Provider Retention Rates. Collecting child-care data is challenging for many reasons. It is often difficult, for example, to keep centers and homes in the sample. Eight centers and six family homes dropped out of the study before or during the Wave 2 data collection. This meant that only 88 percent of the centers and 84 percent of the family homes that were assessed in Wave 1 were also part of the Wave 2 data collection. All of the family homes that did not return in Wave 2 went out of business, as did six of the eight centers that did not return in Wave 2. Of the remaining two centers that did not return in Wave 2, one withdrew from the study because the director was unwilling to host the rating team when a visit was scheduled; the other no longer wanted to participate. An additional three family homes withdrew from the study between Waves 2 and 3 because they went out of business. This left 76 percent of the original sample of family homes in the study until the end. Additionally, nine more centers either went out of business (N = 6) or withdrew (N = 3) between Waves 2 and Wave 3. Thus, 17 of the original centers dropped out at some point during the study, leaving 74 percent of the original sample of centers in the study to the end. Providers were not refreshed at any point. (See the discussion in Chapter Five that compares providers that stayed the course with those that dropped out of the study.)

Participating Teachers

Training and education information about all directors and staff working in classrooms of participating centers as well as family homes was collected, unless they worked less than 30 percent of the time that the center was open.

---

[2] At that time, participation in the study involved major incentives for providers that were randomly assigned to the treatment condition ($7,500 per classroom and a coach was provided); these incentives may have contributed to the remarkably high consent rate. Comparison group providers received only a fraction of these incentives. Participation incentives were smaller after the study design changed from one characterized by random assignment to an intervention or to a comparison condition. (In the appendix to this chapter, see Document A.2.1 for a discussion of the original study design, and Table A.2.2 for a description of participation incentives.)

Participating Families and Children[3]

Parent Recruitment. At the outset of Wave 1, data collectors attempted to recruit preschool-age children and their families by attending parent meetings facilitated by Qualistar staff in centers that had agreed to participate in the study. However, very few parents attended; the meetings were dropped. Instead, center directors and family child-care providers were asked to approach parents themselves about participating in this study—a suboptimal approach, as providers might bias their requests, soliciting parents who were more supportive of the provider or who had higher-functioning children. Qualistar met with all center directors and family child-care providers to inform them of the details of the study so that they could convey accurate information to parents. Data collectors also provided parental consent forms and information packages for directors. The parental consent rate (i.e., the percentage of parents who were approached and consented) for those parents for whom we could calculate a consent rate (N = 568) was very high (94 percent). Consent rates for the remainder of Wave 1 parents were not collected. The parental consent rates for refreshing the sample in Waves 2 and 3 were 77 percent and 73 percent, respectively. Parent participation incentives are described in Table A.2.2 in the appendix to this chapter.

Our goal was to assess and retain approximately 12 children per center who were present across the three waves, which would encompass approximately 24 months of data collection. This number was selected in order to have a large enough sample of children present in the centers at all three waves of data collection to enable the longitudinal analyses. To allow for attrition, the goal in Wave 1 was to assess 20 children in a center, ideally in a single preschool classroom. When it was not possible to assess the targeted number of children in a given classroom, additional preschool classrooms in the center were included. In most cases, two to three classrooms were sufficient to find 20 children. In Wave 2, the goal was to assess 15 children per center, and, in Wave 3, the target was 12 children per center. To maximize the number of children who would be available for longitudinal analyses, in Wave 1 the youngest children within the age range of 2.5–5 years in a given center were selected. In subsequent waves, when the number of children available from the previous waves dropped below the target number, the child sample was refreshed to meet our sample size goals. To maximize the number of children who could be linked to the classroom-level process measures, families whose children were cared for in the classroom containing the largest number of evaluation children were approached first. When target numbers were not reached, families from other preschool classrooms in the center were also approached to elicit their participation in the study. For this reason, in some instances, preschool classrooms may have only one child in the evaluation.

Sometimes, family home providers were unable to obtain consent for at least four children between the ages of 2.5 and 5, usually because there were not sufficient children in that age range being cared for in the home. When less than four consents were obtained from a given family child-care home, children who were older than 5 were included in the Pre-K Snapshot assessment (a measure designed to assess the quality of play opportunities that caregivers provide for children; see the discussion of this measure below), as this measure requires

[3] The original study design, which involved random assignment and following children into elementary school, had to be changed early on because of lack of funding. Providers and parents recruited under the original design remained in the study, but new parents were recruited in a slightly different way (all parents were invited to participate). To improve readability, details of the original design were dropped from the text but may be found in the appendix to this chapter (see Document A.2.1).

that four children be assessed from each classroom. Consents were not obtained from the parents of these children, as the Pre-K Snapshot does not require the observer to interact with the children or collect any identifying information. These children are referred to as the "non-evaluation children" in Table 2.1 below, which presents the number of children assessed in each wave.

As shown in Table 2.1, child attrition was very high, with data on less than 10 percent of the children available across all three study waves. We attempted to mitigate the impact of this huge child turnover by refreshing the sample at Waves 2 and 3, as discussed above. However, the high attrition rate had significant implications for the types of analyses possible as well as the conclusions that can be drawn from the data, as discussed in subsequent chapters.

Measures

The study involved several sets of measures, including the Q-QRIS components and a set of criterion measures that assess process quality and child functioning. Each set is described below.

Qualistar Quality Rating and Improvement System Components

The five Q-QRIS components were collected from all classrooms in a given center.

Classroom Learning Environment. The Qualistar QRIS relies on the Early Childhood Environment Rating Scale–Revised to assess the preschool classroom environment in centers. The ECERS-R is a 43-item inventory that provides a global measure of the quality of a child's preschool or kindergarten classroom environment (Harms, Cryer, and Clifford, 1998). Each of the 43 items is scored on a 7-point scale, with the following categories for each indicator: inadequate (1), minimal (3), good (5), and excellent (7). This rating scale evaluates a range of domains, including personal care routines, classroom furnishings, language-reasoning experiences, fine- and gross-motor activities, creative activities, interactions, and adult needs. The adult needs section was not administered because it relies heavily on self-reports, with their

Table 2.1
Number of Children Assessed in Both Centers and Homes, by Wave

Number of Children	Evaluation Children	Non-Evaluation Children
Total		
Wave 1	1,499	31
Wave 2	824	50
Wave 3	648	44
Returning		
Waves 1 and 2	440	8
Waves 2 and 3	265	3
Waves 1 and 3	141	1
Waves 1, 2, and 3	113	1

inherent informant bias. The scores of all 43 items are then averaged to yield a mean classroom score. This score, which ranges from 0 to 7, is referred to as the "ECERS-R Class Mean" in subsequent sections of this report. (See Harms and Clifford, 1980; Harms, Cryer, and Clifford, 1998; and Sakai et al., 2003, for validity and reliability information for the ECERS, the ECERS-R, and their relationship to each other.)

The ECERS-R data were collected in all preschool classrooms in participating child-care centers by staff who had been trained by the Center for Human Investment Policy (CHIP) based at the University of Colorado, Denver.[4] Trainees had to be reliable on three consecutive ECERS-R administrations, with the criterion for reliability being a score of 85 percent or higher on each item.[5]

Ratios. Ratios were collected in different ways over the course of this study. Initially, center staff were asked to note the child-staff ratio in each classroom at four pre-designated time points each day for four weeks, yielding a total of 80 time stamps for full-time programs. However, concerns about reliance on self-reported data in a high-stakes context led to our collecting ratios using sign-in/sign-out data, as described in detail in Chapter Three. The results of these analyses led Qualistar to rely on observers to collect ratio data. (See Table A.2.3 in the appendix to this chapter for ratio data-collection approaches.)

Staff Training and Education. Training and education data were collected by Qualistar staff who went to the child-care sites included in the study to review and code transcripts and training certificates of all teaching staff in evaluation classrooms. The following three variables were examined for the director and classroom staff: formal education, ECE credits, and years of experience. For the director, administrative experience was also examined. (See Documents A.2.2–A.2.4 in the appendix to this chapter for scoring forms by version.)

Parent Involvement. Parent involvement was included in the Q-QRIS because Qualistar believes that parents can contribute to the quality of a program and that high-quality programs involve parents. However, the best way to assess this concept was not entirely clear. After various failed efforts to measure parent involvement (see Chapter Three), we developed a new instrument, the Family Partnership (FP) measure, described in Chapter Three (see Zellman and Perlman, 2006, for details on our analyses of these earlier measures).[6]

The FP measure includes an 18-item Family Form to be completed by parents and a 16-item Documentation Checklist to be completed by providers. (See Documents A.3.4 and A.3.5 in the appendix to Chapter Three for a copy of these forms.)

The Family Form—also referred to as the "FP parent survey" in this report—focuses on the ways in which the program interacts with parents (e.g., opportunities for volunteering) and has facilitated the child's development and the parent's parenting activities. Each of these

---

[4] The ITERS and Family Day Care Rating Scale (FDCRS) were administered where appropriate and used by Qualistar to generate provider-level scores. However, with the exception of Chapter Six, which focuses on family homes, they are not discussed in this report, as our analysis focused on classrooms that served preschool-age children, and these classrooms are assessed using the ECERS-R.

[5] Following the reliability procedures developed by the authors of the ECERS-R, reliability was calculated by comparing the trainee's score with the "consensus" score agreed on by the trainee and the expert. A disagreement of 1 (e.g., if a trainee scored a particular item as a 3 while the consensus score was 4) was considered a match. Re-reliability testing was done during every 10th observation.

[6] This measure was originally called the Family-Provider Partnership measure and was referred to by this name in some earlier documents.

assessments is made on a 5-point scale. Qualistar totaled the points from the parent survey to produce a summary score, identified as the "FP parent points" later in this report.

The Documentation Checklist—also referred to as the "FP provider survey" later in this report—closely follows the Family Form. As with the FP parent survey, Qualistar totaled the points from the FP provider survey to produce a summary score, identified as the "FP provider points" later in this report.

A summary FP score was created by adding the FP parent points and the FP provider points.

Accreditation. The Q-QRIS assigns accreditation 20 percent of the point value of the other Q-QRIS components. The inclusion of accreditation reflected Qualistar's wish to recognize the importance of the self-study process but at the same time acknowledge its expense and the fact that it measures constructs that overlap with the ECERS-R.

Summing Across Components

Qualistar converts scores on each of the Q-QRIS components to a scaled score and then combines these scores to create a single Q-QRIS score (also referred to as a star rating), which can range from 0 to 4. It does this by assigning points to each of the Q-QRIS components. Assignment of points was decided by Qualistar based on extensive experience in child-care programs. Like other QRISs, there were no data available on which to base these decisions (see Zellman and Perlman, forthcoming, for a discussion of scoring decisions in five pioneer QRISs). The cut scores and relative weights of the components changed over the course of the study as Qualistar tried to modify its scoring system to reflect what it was finding as scores came in. For example, Qualistar staff wanted to make sure that a really great provider would receive more points than a good one; concerns about discouraging staff with low levels of formal education from pursuing additional training and education led to changes in points as well. Given these changes in points over time, to maintain comparability across the versions of the Q-QRIS, we almost always use raw scores in our analyses. For example, rather than using the Training and Education summary score, we use three individual training and education items (years of teaching experience, Early Childhood Education credits, and level of formal education) in most of our analyses. (See Chapter Three for a discussion of these analyses.) However, because star ratings are an important Q-QRIS system output, we do examine the relationship of these ratings to key outcomes in Chapter Five.

Changes Made to the Q-QRIS Components Over Time

The Q-QRIS, although seemingly fully developed when the evaluation began, was in fact a work in progress. Early on, RAND and Qualistar agreed that the ultimate goal was to create and test the best Q-QRIS possible, even if it meant that changes in the way that specific components were operationalized over time threatened the study's interpretability. It made no sense, all agreed, to retain that power at the cost of trying to validate a component that early data clearly indicated was not well understood, was producing no variation in ratings, or could be collected more efficiently or reliably another way. This willingness to change measures continued throughout this study, improving the measures but creating analytic challenges as the Q-QRIS was continually refined.

To work efficiently with the Q-QRIS data, those Q-QRIS components that were changed were classified into "versions" in order to identify which Q-QRIS components changed within and/or between waves of data collection and how. In other words, different versions of Q-QRIS

data represent a change in either data-collection methodology and/or the way in which Qualistar calculated the Q-QRIS scale scores. For example, in Wave 1, ratio data were collected in three different ways (version 2, 3, or 4; version 5 ratio data were collected in Waves 2 and 3 only). (Tables A.2.3 and A.2.4, in the appendix to this chapter, display the different versions of ratio and FP data collected; these were the two components whose measurement [and not just data aggregation] changed over the course of the study.)

Measures to assess the learning environment (ECERS-R and FDCRS) and the accreditation data did not change over time and were therefore not categorized into versions. (For those Q-QRIS components that were assigned versions, see Table A.2.5 in the appendix to this chapter, which indicates which versions of data are associated with specific waves of data collection.)

Criterion Measures

A number of criterion measures were employed as part of the Q-QRIS component validation process. Criterion measures include both measures of process quality and of child functioning. The process-quality measures assess children's actual experience in care, including their interactions with peers and caregivers and the activities they engage in. If they measure levels of quality similar to those measured by the Q-QRIS components, this supports the validity of the component measures. Child-functioning measures enable analyses of the effect of quality on child outcomes. If the Q-QRIS components are picking up quality appropriately, they should co-vary with child outcomes. Changes in quality as a result of quality interventions should improve child functioning, according to the QRIS logic model.

Process-Quality Criterion Measures. Process quality refers to the actual processes that take place in child-care settings. Process quality (e.g., the warmth, sensitivity, and stimulation that characterizes child-staff interactions) is thought to drive the quality of children's experiences in these settings and is assumed to be driven by structural characteristics, such as ratios, staff training, and education, although structural measures are considered to be, at best, a crude proxy for process quality (see, e.g., Scarr, Eisenberg, and Deater-Deckert, 1994). In order to examine the validity of the Q-QRIS, it was critical that we examine its relationships to process-quality measures. The process-quality measures were collected in just one preschool classroom; the decision to collect these data from a single classroom was made when anticipated funding did not materialize and we were forced to dramatically alter the study design and significantly reduce the scope of data collection. This decision limited the analyses we could do, as discussed in Chapter Five.

Pre-Kindergarten Snapshot. The complexity of child-child interactions was assessed with the Pre-K Snapshot (Howes, personal communication, 2001). Data from the Pre-K Snapshot were used to assess the type and level of play that children experience in a given classroom. The Pre-K Snapshot contains several different scales and dimensions, including nonplay activity type, peer play, cognitive play, and adult involvement. (See Document A.2.5 in the appendix to this chapter for the Pre-K Snapshot Score Sheet.) Portions of the scale have been used in many different studies (Howes, 1997; Howes and Matheson, 1992; Howes, James, and Ritchie, 2003; Howes and Smith, 1995; Howes, Smith, and Galinsky, 1995; Howes and Stewart, 1987; and Ritchie and Howes, 2003). (See Tonyan and Howes, 2003; and Wishard et al., 2003, for reliability and validity information.)

Four children in each assessed classroom were randomly selected and systematically observed 50 times each over the course of three to four hours. Because the children were ran-

domly selected and their data subsequently averaged into a classroom-level score, the type of play that they engage in was assumed to represent the level of play of all children in the classroom. The types of play assessed include nonplay, peer (social) play, and cognitive play. The levels of play range from solitary play (playing by oneself) to complex pretend play (when the target child and at least one peer are engaged in social interaction in which there is evidence of a "script" and each child plays a given role). Adult involvement (or lack of involvement) was noted whenever an adult was within three feet of a target child. The 50 observations for each child were summarized to yield a score that indicated the proportion of the observed time during which the target behavior was evident.

Reliability training for the Pre-K Snapshot was conducted by UCLA staff for staff of the OMNI Institute, a Denver-based research institute, which collected the data. With one exception, data collectors obtained a 70 percent agreement or higher on all Pre-K Snapshot categories.

Caregiver Interaction Scale (CIS). The CIS (Arnett, 1989) is a 26-item classroom-level measure that assesses the quality and content of teachers' interactions with their students. The scale provides information on socialization practices, with items tapping the emotional tone, discipline style, and responsiveness of caregivers. The items are organized into four subscales: (1) positive relationship (warm, enthusiastic, and developmentally appropriate behavior), (2) punitiveness (hostility, harshness, and use of threats), (3) detachment (lack of involvement or apparent interest), and (4) permissiveness. (See Document A.2.6 in the appendix to this chapter for the CIS.) Layzer, Goodson, and Moss (1993) examined the concurrent validity of the CIS and found correlation coefficients of .43 to .67 between the CIS and the ECERS and between the CIS and two other measures, the Assessment Profile for Early Childhood Programs and the Description of Preschool Practices. The authors did not expect the coefficients to be large because the CIS focuses more narrowly on aspects of teacher behavior not directly measured by the other three observation instruments. However, Phillipsen, Cryer, and Howes (1995) reported a correlation of .76 between the CIS and the ECERS.

The CIS was completed at the end of the Pre-K Snapshot observations, because after four hours of intense observation of the teachers' interactions with children, observers had the information they needed to make informed judgments about the teachers' behavior (the test developer suggests that the CIS can be completed based on a minimum of 45 minutes of observations). The CIS was always completed by data-collection staff with the head teacher in mind. When it was not possible to complete the CIS on the head teacher, the teacher acting as the head teacher in the classroom on the day of assessment was rated instead. An initial group of raters was trained by Carollee Howes to administer the CIS. (See Table A.2.6 in the appendix to this chapter for response rates for the Pre-K Snapshot and CIS in the home sample.)

Child Outcome Criterion Measures. As shown in the logic model presented in Chapter One, QRISs are posited to improve child outcomes as programs improve the quality of care. Hence, part of the validation effort focused on child outcomes. Since the intervention was designed to improve quality, the hypothesis was that this improved quality, if it occurred, would, over time, improve child outcomes for children who attended frequently enough and stayed long enough that the quality of the provider could affect their functioning. We were interested in a comprehensive view of children's functioning and thus considered their well-being broadly. For this reason, measures of both social and cognitive development were selected. To obtain comprehensive data about each child, we collected information about chil-

dren's functioning from multiple informants using multiple data-collection methods, including surveys from parents and teachers and direct assessments of children.

To assess child cognitive functioning, we chose the Peabody Picture Vocabulary Test–III and three subsets of the Woodcock-Johnson achievement test (Letter-Word Identification, Passage Comprehension, and Applied Problems), discussed below. Together, these measures represent a range of key cognitive skills that have been associated with school readiness. Data-collector training on these instruments was conducted by a facilitator with extensive experience in child development and standardized testing.

The Peabody Picture Vocabulary Test–Third Edition (PPVT-III). The PPVT-III, a standardized test of receptive language skills (Dunn and Dunn, 1997), was selected because of its wide use in clinical and research studies. It also poses a minimal burden on children as the test quickly focuses on items at their current level of functioning, reducing frustration and assessment time. The PPVT-III was administered to all consented children. The test consists of 204 questions grouped into 17 sets of 12 items that are arranged in order of increasing difficulty. For each item, children are presented with four black-and-white pictures and asked to select the picture that best represents an orally presented word. The starting point of the test is determined by the child's age. Item sets that are too easy or too difficult for a child (as dictated by a standardized administration protocol) are not administered. The test yields one overall standardized summary score representing the child's level of receptive language skills—how well a child can understand what is said. (See Dunn and Dunn, 1997, for reliability and validity information.)

Woodcock-Johnson–Third Edition, Tests of Achievement (WJ-III). Three subtests of the WJ-III were administered to consented children (Woodcock and Johnson, 1990). To evaluate children's pre-verbal skills, the Letter-Word Identification and Passage Comprehension subtests were administered. The Letter-Word Identification task tests children's symbolic learning (i.e., matching pictures with words) and reading identification skills (i.e., identifying letters and words), while the Passage Comprehension task measures children's comprehension and vocabulary skills using both a multiple-choice and fill-in-the-blank format. Additionally, children's pre-math skills were evaluated using the Applied Problems subtest of the WJ-III, a measure of children's ability to solve practical mathematics problems. Three standardized summary scores were computed for each participating child. (See Woodcock et al., 2006, for reliability and validity information.)

We augmented the cognitive measures with the Child Behavior Inventory (CBI) and the Strength and Difficulties Questionnaire (in Wave 3 only). Each is described in turn below.

The Child Behavior Inventory. The CBI is a 60-item questionnaire that is administered to teachers to assess children's ability to adapt to and participate in the preschool environment (Schaefer, Edgerton, and Aaronson, 1978). Teachers are asked to rate how well each item describes a particular child on a 5-point scale ranging from "not at all" (1) to "very much" (5). Factors examined in the CBI include verbal intelligence, introversion/extroversion, dependence/independence, considerateness, apathy, task orientation, hostility, creativity/curiosity, and distractibility. The introversion/extroversion subscales were excluded from the analyses because they are challenging to interpret and culturally loaded. (See Document A.2.7 in the appendix to this chapter for this instrument.)

Mirante and Ryckman (1974) and Blunden, Spring, and Greenberg (1974) used factor-analytic methods to assess the structure of the CBI. Both studies found that the CBI is stable in

terms of factor structure across different age groupings and geographical areas, and both found that it measured three dimensions of behavior. However, the dimensions varied slightly.

Strength and Difficulties Questionnaire (SDQ). The SDQ is a brief behavioral screening questionnaire that asks parents to describe 25 attributes of the participating child, some positive and some negative. These 25 items are rated on a 3-point scale: "not true," "somewhat true," and "certainly true." Items are divided among five scales: (1) emotional symptoms, (2) conduct problems, (3) hyperactivity/inattention, (4) peer relationship problems, and (5) prosocial behavior. (See Document A.2.8 in the appendix to this chapter for this questionnaire.)

Family and Teacher Background Measures

Family Background Questionnaire (FBQ). It is imperative to control for family background characteristics when using child-functioning data. For this reason, family background information was collected. The Family Background Questionnaire included 20 items assessing the family's socioeconomic status, education level, and ethnicity; 11 items targeted at evaluating the child's child-care history (e.g., average hours per week the child attends the provider, average months the child has attended the provider), five items that determine the child's current level of health (e.g., how often does the child get sick or injured at the child-care center), and several items that are specific to the child (e.g., child has learning problems, age at assessment). (See Documents A.2.9–A.2.11 in the appendix to this chapter for these measures.) Parenting style was captured by the Block Child-Rearing Practices Report (Block, 1965, 1980). This is a self-reported measure that captures parents' attitudes toward parenting. It consists of a series of items rated on a 6-point scale (from 1 = "not at all descriptive of me" to 6 = "highly descriptive of me"). Two subscales, nurturance and restrictiveness, were included. (See Document A.2.12 for the Child-Rearing Practices Report.)

Despite incentives, only 64 percent of parents returned the 104-item Family Background Questionnaire at Wave 1. Many FBQs included missing responses (parents were told to skip items they did not wish to answer). This led us to shorten the FBQ in Wave 2 to 74 items by omitting 28 child behavior items and two particularly sensitive items on breastfeeding and religiosity. Wave 2 response rates continued to be low (60 percent). In Wave 3, we added the 25-item Strengths and Difficulties Questionnaire in order to include the parents' perception of the participating child's behavior as well as five additional questions targeting the child's level of health. In an attempt to improve the response rates in Wave 3, we increased the completion incentives, but Wave 3 response rates remained at 60 percent.

Teacher Questionnaire. The Teacher Questionnaire was a short survey completed by the head teacher in each class; six items were directed at determining his or her demographic and education background, teaching experience, and job satisfaction. Extensive training and education information was collected separately for all staff in the center as part of the Q-QRIS, as described above. (See Document A.2.13 in the appendix to this chapter for this questionnaire.)

Table 2.2 displays the response rates for the Q-QRIS component measures, the process-quality measures, the child-functioning measures, and the family characteristics measures in centers. (See Table A.2.6 in the appendix to this chapter for analogous data for the family child-care homes sample.)

Table 2.2
Response Rates for Criterion and Q-QRIS Component Measures (Centers)

		Number of Responses (N) and Response Rates (%)					
		Wave 1		Wave 2[a]		Wave 3	
Domain	Measure	N	%	N	%	N	%
QRIS components	FP parent survey	2,057	X[b]	1,914	X[b]	2,121	X[b]
	FP provider survey	62	95	49	94	42	86
	ECERS-R class mean	156	100[c]	129	100[c]	109	98[c]
	Average class ratio	161	99	132	100	109	98
	Director A.A. degree	59	91[d]	52	94[d]	42	86[d]
	Director B.A. degree	59	91	52	94	42	86
	Director administrative experience	59	91	7	13	42	86
	Accreditation status	65	100[e]	52	100[e]	49	100[e]
	Head teacher A.A. degree	124	76	100	75	87	78
	Head teacher ECE credits	124	76	100	75	87	78
	Head teacher experience	124	76	100	75	87	78
Classroom Ns		156		132		111	
Provider Ns		65		58		49	
Process-quality measures	Pre-K Snapshot	75	100[f]	54	100[f]	51	100[f]
	Caregiver Interaction Scale	75	100[g]	54	100[g]	51	100[g]
Child functioning	Peabody Picture Vocabulary Test	1,184	87	709	86	576	92
	Woodcock-Johnson	1,232	90	685	83	582	93
	Child Behavior Inventory	1,320	96	742	90	532	85
	Strength and Difficulties Questionnaire	N/A	N/A	N/A	N/A	334	54
Child Ns		1,368		774		558	
Family characteristics/ covariates	Demographic Survey	871	64	464	56	334	54
	Child-Rearing Practices Report	871	64	464	56	334	54
Family Ns		1,360		829		619	

NOTES: A.A. = associate of arts degree; B.A. = bachelor of arts degree; N/A = not applicable.

[a] Because of logistical problems, including a change in study design (described in Document A.2.1 in the appendix to this chapter), the lag time between Wave 1 and projected Wave 2 data collection became so great in some cases that some providers had to be dropped from the Wave 2 data collection altogether. Six centers and seven family home providers were lost at Wave 2 because of this problem. Because they were dropped from Wave 2 data collection because of factors independent of their own characteristics, they were not included in the calculation of Wave 2 response rates.

[b] The number of FP Parent Surveys that were distributed to center parents is unknown.

[c] The age composition of the children in each classroom determined whether the ECERS-R (appropriate for classrooms serving preschool-age children) or ITERS (appropriate for infants and toddlers) was administered. Some preschool-age children included in the study were in classes in which the ITERS was administered rather than the ECERS-R; these children did not have ECERS-R scores associated with their other data. While their ECERS-R scores were not missing in the strict sense because the ECERS-R was not administered in their classroom, they were missing for analytic purposes, and their scores were therefore imputed. In contrast, ratios were collected from all rooms. This explains the discrepancy between the number of classrooms with ratios and with ECERS-R scores.

[d] Qualistar staff collected training and education information from all staff in evaluation classrooms.

[e] Qualistar staff verified whether all centers were accredited.

[f] The Pre-K Snapshot was collected in one evaluation class per center.

[g] The CIS was collected in one evaluation class per center.

Organizational Change Survey

The Organizational Change Survey (OCS) was administered across all three waves of data to provide some information on provider operations, e.g., frequency of staff meetings, and some detail on the changes providers decided to implement in response to their quality-feedback sessions. This was the only information available on Q-QRIS implementation and was not part of the quality rating system. In Wave 1, both staff and directors filled out OCSs. Wave 1 analyses indicated that director and staff responses were highly correlated, so the staff survey was dropped in Waves 2 and 3. The survey included 32 items that asked about key implementation activities, such as the frequency and content of staff meetings, planning time, child-care fees, staffing plans, professional development opportunities, goal-setting priorities, staff movement among classrooms, staff turnover, training opportunities and support, and an assessment of current quality and plans to improve. The response rate for the OCS was 90 percent at Wave 1 (45 of the 50 centers that were asked to complete a survey returned one); 73 percent at Wave 2 (38 out of 52), and 77 percent in Wave 3 (37 out of 48).

Data-Collection Procedures: Providers, Parents, Staff, and Children

Data Collection

Qualistar staff collected all Q-QRIS data. Criterion data were collected by the OMNI Institute. Center directors and family home providers documented their Family Partnership activities, another Q-QRIS component, for Qualistar staff to review.

QRIS Data-Collection Protocol

Child-care programs were assigned an annual one-month rating window. One month before the rating window, programs submitted background information about the site, including names of staff, names of classrooms and age groups served, profit status, Head Start status, hours of operation, type of site, curricula used, and the program's funding sources.

During the one-month rating window, Qualistar staff conducted an unannounced classroom-environment observation (ECERS-R, ITERS, or FDCRS, depending on the type of program and age groups served in the classroom). Ratio data were collected during the rating period in a number of different ways, as discussed above. Center directors were also given training and education forms to fill out for each staff member several weeks before the rating window to ensure that they had enough time to collect all necessary documentation. Directors reported on the years of experience that each staff member had in either teaching or administrative positions. Training certificates dating back three years and transcripts were collected for each staff member working in the program for at least 30 percent of the time that the program was open. Qualistar staff reviewed the transcripts and training certificates and assigned the appropriate points to each staff member.

Family Partnership surveys were provided to each participating program to distribute to each participating family several weeks before their rating window. Qualistar collected the completed surveys from the program during the rating period. Families returned the surveys either to the center director or the classroom teacher in privacy envelopes. Qualistar also interviewed program administrators regarding their FP activities and asked programs for supporting documentation (such as family night flyers, meeting minutes, and parent involvement

plans). Programs were given the FP assessment criteria several weeks prior to the rating window so that they would have time to collect all supporting documentation.

If the program was accredited, Qualistar collected a copy of its accreditation report and certificate.[7]

Once all of the data were collected during a program's one-month rating window, Qualistar staff manually calculated a rating and entered the data into a database that also calculated the rating. If any discrepancies were found between the manual calculation and the computer-generated calculation, the discrepancy was located and the source assessed, then the inconsistency was resolved.

Qualistar staff used rating results from each component to write a quality-performance profile for the program. This profile detailed the program's strengths and areas for improvement for each rating domain. It identified trends across the program as well as items that were specific to each classroom. The profile also included a quality-improvement plan that identified strategies to assist programs and classrooms in improving their practice as well as resources needed and a timeline for improvement. The rating calculation and data-entry process and the production of the quality-performance profile were scheduled to take approximately a month. This rapid turnaround was intended to allow programs the maximum amount of time to improve quality before another assessment. However, during the first year of this study, this process generally took considerably longer, sometimes up to four months. During the second and third years of this study, programs received their reports in a timely manner: one month after their rating window ended.

Qualistar staff consulted with each program to outline the results of the rating and to discuss ideas about how to implement the quality-improvement plan. Programs were given the opportunity to contest their rating within a one-month time period following this consultation. In general, programs that disputed their rating had failed to submit transcripts for all staff in the program. Additional transcripts were then reviewed and additional points awarded if appropriate.

Criterion Data-Collection Protocol

OMNI Institute staff distributed the key surveys that constituted some of the criterion measures, such as the Classroom Behavior Inventory and Teacher Questionnaire forms, to the head teacher in each classroom. When classrooms had no head teacher, OMNI Institute staff gave the questionnaires to the primary caregiver(s) in the classroom. Teachers were not assigned identification numbers. As a result, there is no way of knowing who actually filled out the forms. However, OMNI Institute staff were told that the child surveys were often divided between the head teacher and assistant teacher.

Each assenting child was assessed by a trained OMNI Institute data collector for approximately 30 minutes. During this time, children were administered the three subtests of the WJ-III and the PPVT-III. Children were removed from their peer group for the assessment and were taken to a quiet corner of the classroom or to a table set up in the hall or in a vacant

[7] Qualistar recognized accreditation by the National Association for the Education of Young Children, the National Association of Family Child Care, American Christian Schools International, and the Council for Exceptional Children. If a program was accredited by any other agency, the director was asked to submit the accreditation standards to Qualistar for review. This happened in only one instance, and in this instance, Qualistar reviewed the standards and considered the program accredited.

office or classroom to minimize distractions. Questions that children refused to answer were skipped. Children were never informed about the correctness of their responses, but were praised for their participation.

When OMNI Institute received parental consent forms from participating child-care providers, their staff either sent Family Background Questionnaires to classroom teachers who would hand them out to parents, or placed FBQs directly in the target child's classroom cubbyhole. Parents were asked to return the completed FBQ directly to OMNI Institute in the postage-paid envelope provided. All versions of the FBQ were made available in Spanish as well as English.

Staff were asked to complete a Child Behavior Inventory for each consented child. These two surveys took approximately 20 minutes to complete per child. Teachers received a stipend of $3.50 for each child that they assessed. Teachers were also asked to fill out the Teacher Questionnaire, which asked about their ethnicity and teaching experience. All teacher surveys were completed in English. OMNI Institute left these forms in classrooms, ideally with the head teacher; when this was not possible, forms were left with the primary caregiver in the classroom at the time of the assessment visit. OMNI Institute staff collected the forms on the last day of assessment in the center. If teachers were not finished with the questionnaires by that time, data collectors either returned to the center to collect the remaining surveys at a later date, or they provided the teachers with a postage-paid envelope in which to return them.

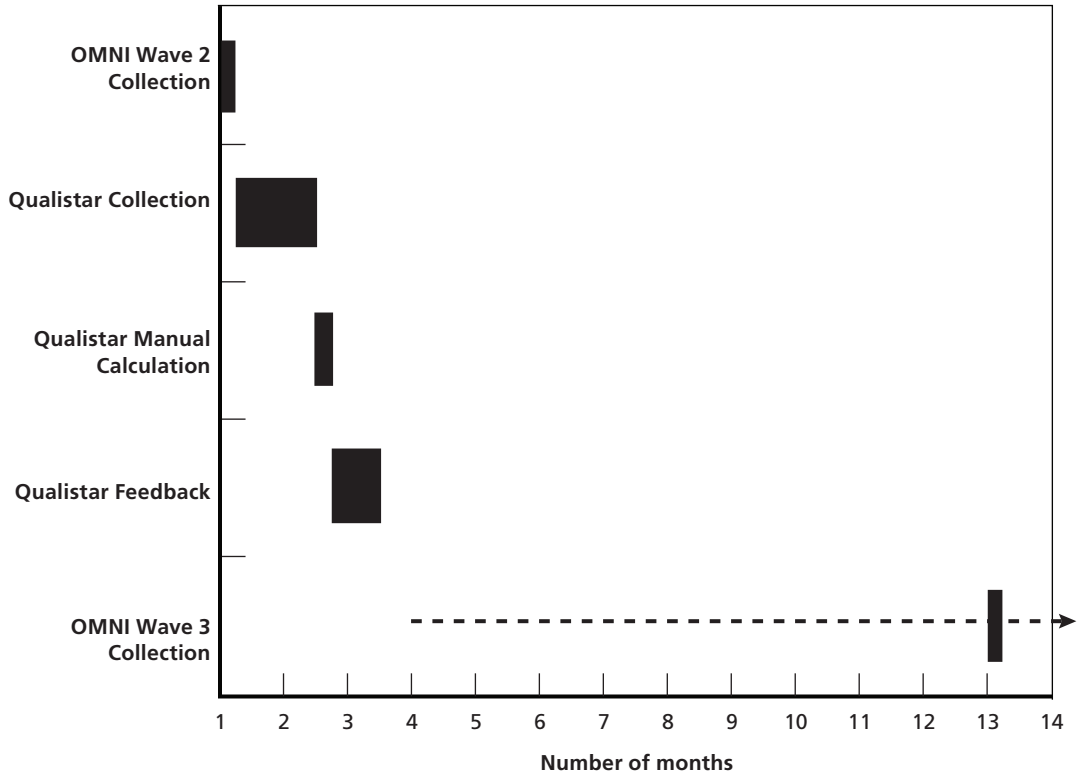
Data-Collection Timeline

The initial project schedule specified three waves of data collection, occurring at approximately 12-month intervals. In each wave, OMNI Institute would carry out their assessments before Qualistar administered the Q-QRIS. The proposed timeline is presented in Figure 2.1.

Early in 2004, when the Wave 1 data collection was ending and OMNI Institute was preparing the Wave 2 data collection, it became apparent that the time between Wave 1 and Wave 2 assessments varied greatly among the providers because of lags in assessment and feedback. Some providers had a very short window of time in which to make the changes that were recommended in their feedback sessions, while other providers had much more time. The total time from the point at which OMNI Institute first entered a center, Qualistar conducted its rating, scores were entered and calculated, and the site received feedback was anticipated to be between eight and nine weeks, as shown in Figure 2.1. But Table 2.3 below shows that the elapsed time during Wave 1 from the start of OMNI Institute's data collection to Qualistar's debriefing session was often far longer. Indeed, in 70 percent of the sites (35 out of 50), feedback on the Wave 1 assessment was provided more than 10 weeks after OMNI Institute had finished its data collection, and in some sites the elapsed time was more than 20 weeks.

Because of logistical problems, including a change in study design (described in Document A.2.1 in the appendix to this chapter), the lag time between Wave 1 and projected Wave 2 data collection became so great in some cases that some providers had to be dropped from the Wave 2 data collection altogether. Six centers and seven family home providers were lost at Wave 2 because of this problem. However, these providers were included in the Wave 3 data collection, so they were not included in the attrition calculations. And because they were dropped from Wave 2 data collection because of factors independent of their own characteristics, they were also not included in the calculation of Wave 2 response rates.

Figure 2.1
Evaluation Activity Timeline per Center



RAND *MG650-2.1*

Table 2.3
Time Lag Between OMNI Institute Wave 1 Collection and Qualistar Feedback

	Days Between OMNI Collection and Qualistar Feedback								
	0–28	29–42	43–56	57–70	71–84	85–98	99–112	113–126	127–154
No. of providers	1	3	7	4	5	4	14	10	2

Analytic Approaches

Small sample sizes available for analyses influenced the statistical models and techniques we used. The fact that fewer than 10 percent of the children in the original Wave 1 sample were present at Wave 3 meant that a longitudinal analysis modeling growth in children's outcomes over the three waves was not possible. Consequently, we limited our statistical approaches to cross-sectional analyses and analyses that examined the relationships among outcomes and quality indicators over two consecutive waves. (More information about these analytic approaches is discussed below.)

Lack of parental response to the FBQ also contributed to small sample sizes. Across waves, approximately 50 percent of children who had been assessed did not have data relating to income, other demographic information, and parents' child-rearing style. However, children whose parents returned the FBQ performed similarly to children with missing demographic

information, suggesting that missing data may be random. That is, there does not appear to be a bias in the sample with respect to missing background data.

The decision to assess process quality from only one classroom per center (and that classroom changed each year in order to capture as many evaluation children as possible) meant that the analysis could only be conducted on a subset of the sample, namely those children who were in the classroom in which the process-quality measures were assessed. Similarly, any attempts to investigate teacher characteristics in relation to process quality were limited to the few teachers whose classrooms were assessed. This decision, made in response to the funding shortfalls that necessitated a change in the study design, sharply limited analytic power. At Wave 1, which was the wave in which the sample size was the largest, process-quality data were available in only 65 classrooms, and this number decreased as data collection went forward.

Finally, as discussed earlier, the evolution of the Q-QRIS over time contributed to the small sample sizes available for analysis. Changes in the FP component in particular led to a reduction in the sample size, especially in Wave 1. In Wave 1, two different versions of the FP were administered, and the measures assessed such different constructs that it was not possible to include both versions in the same analysis without compromising the interpretability of the FP measure. Moreover, changes in the FP measure from one version to the next meant that some providers were not administered the FP during the transition period. This further reduced the sample.

A consequence of the attrition in participating children over time, nonresponse from parents on FBQ, and the changing Q-QRIS measures was a small, potentially nonrepresentative sample size with which to conduct analyses. Because small sample sizes limit the magnitudes of effects we can observe, we attempted to maximize our sample size by undertaking multiple imputation techniques, described below.

Multiple Imputation

To address the large amount of missing data, we implemented a multiple imputation technique in which each missing value was replaced with a random sample of plausible variables. For child-level variables that were not expected to change over time (e.g., gender or race), the imputation occurred at the level of the child. For child-level variables that could be expected to change over time (e.g., hours of care), we imputed at the level of child by wave using all observed data in a multivariate normal model as a basis for predicting missing values. Analogous imputation processes were conducted for missing classroom- and provider-level variables. Dichotomous variables were treated as continuous in the imputation process, then rounded back to dichotomous variables for analysis. We imputed 10 sets of plausible values, then synthesized the results in a manner that accounted for the uncertainty due to the missing information (see Schafer, 1997).

However, some of the analyses described in Chapter Three are not based on imputed data; use of non-imputed data is noted in every instance.

Multiple Statistical Testing

Our comprehensive approach to analyses led us to conduct over 5,000 hypothesis tests. The number of statistical tests we conducted runs the risk of an inflated Type II error (i.e., rejecting the null hypothesis when the null hypothesis should not have been rejected). To minimize the Type II error, we used the False Discovery Rate (FDR) procedure (Benjamini and Hochberg, 1995), in which we adjusted the p-value significance criterion from the traditional .05 to .025.

That is, we considered p-values less than or equal to .025 to be significant. This makes it more likely that any significant results observed are reflective of true effects as opposed to chance.

In some instances, we also discuss marginally significant results, which are findings with p-values greater than .025 but less than .05. We limit these discussions to analyses involving family homes, because sample sizes did not exceed more than 38 providers.

CHAPTER THREE

Analyses of Q-QRIS Components

Since RAND was not involved in the development of the original Q-QRIS and the selection of the component measures, our first step was to examine each of the Q-QRIS components. In this chapter, we describe our efforts to analyze and, in some instances, improve on the five Q-QRIS components: classroom environment; ratios; staff education, training, and experience; family partnerships; and accreditation. Where possible, we drew on existing data to answer basic questions about each construct. Because a number of these analyses rely on different databases from those presented in Chapter Two, each section includes—in addition to background, findings, and conclusions—a brief methods section.

Classroom Environment[1]

Background

Comprehensiveness and the absence of other options have made the Early Childhood Environment Rating Scale (Harms and Clifford, 1980) and its associated measures, the Infant Toddler Environment Rating Scale (Harms, Cryer, and Clifford, 1990) and the Family Day Care Rating Scale (FDCRS; Harms and Clifford, 1989), among the most widely used measures of the quality of the classroom environment. The ECERS was first developed in 1980 by Harms and Clifford and included 37 items grouped into seven subscales. Items were drawn from research, from performance indicators of quality child-care and early childhood programs, and from nominations by child-care practitioners. The authors tested the validity of the ECERS by asking child-care experts to rate the importance of each scale item to early childhood programs. The scale was then tested by comparing its ability to distinguish between classrooms of varying quality as determined by expert observers who had been working with the staff in those classrooms. When ratings on the scale made by expert observers were compared with the trainers' ratings on 18 classrooms, a rank order correlation of .737 was obtained.

A 1998 revision updated and expanded the ECERS, now 43 items, to reflect changes in the early childhood field; items that address issues surrounding children with disabilities and that measure cultural sensitivity in materials and activities were added. The ECERS-R maintains the same conceptual framework, scoring approach, and administration as the ECERS. The authors argue that the revised edition is equally valid (Harms, Cryer, and Clifford, 1998). Indeed, when ECERS-R ratings were compared with ECERS ratings of the same classroom,

[1] This work is a summary of Perlman, Zellman, and Le (2004).

the results suggested that the original ECERS and the ECERS-R can be viewed, as their authors intended, as comparable measures of quality.

Qualistar chose to include the ECERS-R in its Q-QRIS given that there were few alternative measures of equal comprehensiveness. Indeed, the ECERS-R and its related measures are regarded as the "gold standard" in the field. A number of states have included the ECERS-R in their quality rating systems.[2]

But despite its widespread use over many years, little attention has been paid to the psychometric properties of the ECERS or ECERS-R. Indeed, Paget (2001) noted, "certainly, a measurement tool that contributes so significantly to the quality of programs for young children deserves more sustained empirical support." Those studies that have examined the ECERS have not replicated the subscales through factor analysis (e.g., Munton et al., 1997; Rossbach, 1990). For example, Munton et al. (1997) found only one ECERS component in their analysis of data from 113 classrooms. In their analysis of data from 191 classrooms, Tietze et al. (1998) found that the intercorrelations among the subscales were quite high and that there were not seven but instead only two factors. Their cross-cultural data suggest that the ECERS can best be presented in two factors, educational interaction and educational space and materials.

Nor has much attention been focused on the conceptual attributes of the ECERS or the ECERS-R. Both measures clearly set out to measure multiple aspects of the early childhood environment. While these include both aspects of the physical environment and more process-focused aspects of the environment, such as caregiver-child interactions, the instruments clearly focus on the former. Many of the ECERS and ECERS-R items rate aspects of the physical environment and safety. Furthermore, some of the more important interactive measures are self-reported, with the inherent bias that self-reports pose, particularly in the high-stakes environments in which QRISs operate. For example, in assessing greeting and departing on the ECERS-R, providers are asked, "Can you describe what happens each day when the children and parents arrive and leave?" It would be especially important to determine whether these self-reported items, which are more susceptible to bias but are easier to collect, produce ratings that are equivalent to those produced by the observational items.

The lack of attention to these measurement issues has become more relevant of late because the ECERS-R is increasingly being employed in high-stakes settings in which a facility's ECERS-R score could increase or decrease the amount of public funding it receives. Research on high-stakes testing in school settings (Corbett and Wilson, 1991; Shepard and Dougherty, 1991) generally finds that as classroom teachers begin to understand these new high-stakes systems, they rely less on curriculum and curriculum standards as a basis for instruction and increasingly focus their instruction on those skills and concepts that are tested. There is no reason to think that child-care providers would respond differently than elementary and secondary school teachers to such high-stakes contexts. If, as has been proposed, salary and bonuses for child-care teachers are tied to ECERS-R scores, motivation to produce higher ECERS-R scores is likely to increase. This makes the psychometric properties of the ECERS-R particularly important.

A significant exception to the dearth of empirical attention and the use of small samples in studies of the ECERS may be found in the work of Scarr, Eisenberg, and Deater-Deckard (1994). These authors examined the psychometric properties of the ECERS and ITERS through

[2] But some states have decided not to use it as part of QRIS ratings because it is quite costly to administer (see Zellman and Perlman, forthcoming, for a discussion of other ways in which five pioneering states use the ECERS-R).

analysis of 363 classrooms in 120 child-care centers in three states in a low-stakes setting. They found no empirical support for the seven a priori ECERS subscales. Instead, they found that only one factor emerged from a factor analysis of the subscales. This finding held regardless of the instrument, age group, or type of factor analysis.

As part of our effort to validate Qualistar's multicomponent Q-QRIS, we set out to examine how well the ECERS-R appeared to be measuring the classroom environment.[3] Our analyses focused on addressing the following questions:

1. How closely related are the individual ECERS-R items?
2. How many distinct aspects of quality are actually measured by the ECERS-R?

To do this, we used readily available data, some of which were collected as part of this study; the rest were collected by the Center for Human Investment Policy at the University of Colorado, Denver, under a contract with Qualistar. Our goals were to assess the ECERS-R's reliability and to examine its validity as a measure of classroom quality. By assessing key psychometric properties of the ECERS-R, we and others would acquire information about its reliability and other properties.

Methods

Data Collection. ECERS-R data were collected from 326 classrooms in 202 Colorado child-care centers by 41 data collectors over the period from fall 2000 to summer 2002 by staff of the CHIP at the University of Colorado, Denver, and Qualistar. All data collectors met the reliability criteria established by the developers of the ECERS-R. Reliability was established after initial training and rechecked after every 10th administration of the instrument to avoid observer drift. We also collected data about staff credentials and child-staff ratios from the 66 classrooms in 35 providers that were part of this study at that time.[4]

Analytic Approach. We correlated the individual items with each other and with the total ECERS-R score. We also conducted an oblique factor analysis. We used data from the 35 providers from which we had collected information about staff credentials and ratios, and we correlated the ECERS-R score with these measures.

Findings

The average ECERS-R score for the entire sample was 5.153,[5] which is considered very good, using a classification system developed by Howes, Phillips, and Whitebook (1992). This relatively high average likely reflects the fact that all providers in this study were participating in quality-improvement initiatives. Many had previously been assessed using the ECERS-R, and some had received quality-improvement support based on previous ECERS-R performance.

For each ECERS-R item, we assessed the distribution and eliminated those questions that lacked variance (i.e., less than 0.10), showed skewed distributions, or had nonresponse rates above 10 percent. This led us to delete a total of seven items pertaining to meals and snacks;

---

[3] In other analyses reported in Perlman, Zellman, and Le (2004), we also examined whether this concept could be measured more efficiently by using subsets of items.

[4] Data-collection procedures are described in Chapter Two.

[5] The average ECERS-R for the subsample of centers for which we had data on staff credentials and child-staff ratios was 4.23 (standard deviation = 1.64).

naps and rest; safety practices; provisions for children with disabilities; use of TV, videos, and computers; adequacy of space for gross motor movement; and staff cooperation.

We expected the seven ECERS-R subscales to be related to each other to some degree, since they all assess aspects of quality, and different quality components tend to be found together (e.g., Phillips and Howes, 1987). However, if the ECERS-R measures seven *distinct* aspects of quality, the correlations among these different subscales should be of only moderate magnitude: large enough to account for the likely interrelatedness of different dimensions of child-care quality, but not so large that the subscales are essentially measuring the same construct.

We first examined the correlations among the individual items. The average inter-item correlation was .39, and item-total correlations ranged from .35 to .76, with a median of .63. The magnitude of these correlations is fairly high, especially considering that individual items are not as reliable as composites or subscales (Light, Singer, and Willett, 1990). Subscale correlations were also moderately high, ranging from .48 to .76, with a median value of .62. The internal consistency estimate, which is an indicator of the extent to which the items measure the same construct, was .95. These statistics are similar to those found by Scarr, Eisenberg, and Deater-Deckard (1994) and suggest that the ECERS-R does not measure seven distinct aspects of quality.

The factor analysis results confirmed that the items are highly correlated, with the common variance constituting approximately 55 percent of the total variance. Three factors were retained by the Kaiser criterion. The first factor had an eigenvalue of 13.85 and explained 71 percent of the common variance. Items that tap into child activities, program structure, and space and furnishings loaded on this first factor. The second factor had an eigenvalue of 1.93 and explained 10 percent of the common variance. This second factor included child-staff interactions, including personal care routines and the encouragement of language development. The third factor had an eigenvalue of 1.12 and explained 6 percent of the common variance. Items about provisions for parents and staff loaded on this final factor. (See Perlman, Zellman, and Le, 2004—Document A.3.1 in the appendix to this chapter—for a list of the ECERS-R items that loaded onto each of the factors.)

These findings indicate, despite the emergence of these three factors, that the ECERS-R is substantively a unidimensional measure of quality. The fact that the eigenvalue of the first factor is seven times larger than that of the second factor indicates the presence of a single dominant factor. Furthermore, the second and third factors correlated highly with each other as well as with the first factor, suggesting that all three factors assess similar aspects of quality. In our sample, the intraclass correlation, which is a measure of the degree to which the variance can be attributed to cluster (provider) membership, is very high at .73. This indicates that there is much consistency in ECERS-R scores across classrooms of the same provider. These results remained virtually unchanged when, to control for nesting of classrooms within providers, we averaged ECERS-R scores across classrooms within providers and conducted a provider-level factor analysis. This finding may hold promise as a way to reduce rating costs (by reducing the number of classrooms that are ECERS-rated). More work should be done to determine whether these findings hold more generally and across classrooms serving children of different ages.

When we assessed the relationships between the full ECERS-R and staff credentials and ratios, we found that there was a .41 correlation between the complete ECERS-R measure and

total years of teaching experience. The ECERS-R correlated only with caregivers' year of experience, not their educational attainment, post-secondary ECE credits, or child-staff ratios.

Discussion

Our findings for the ECERS-R replicate Scarr, Eisenberg, and Deater-Decker's (1994) analyses of the original ECERS in most respects. We found that the ECERS-R does not assess seven distinct aspects of quality, but instead appears substantively to be a unidimensional quality indicator. As such, the interrelationships of the ECERS-R items do not support the seven-subscale framework posited by its developers. The finding that the ECERS-R is essentially a unidimensional measure of classroom quality suggests that using the total score, as is done in the Q-QRIS, is appropriate. However, our analyses, based on data largely collected in low-stakes contexts, do not address the validity of the ECERS-R in the high-stakes contexts most often found when QRISs are used; validation needs to happen within the context in which the measure is used (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). To date, much of the research using the ECERS-R has been conducted in low-stakes contexts. However, the ECERS-R is currently a part of the Q-QRIS. Consequently, we included it in the regression models described in Chapters Four and Five, and we assess the extent to which the ECERS-R is related to other measures of classroom quality and to child functioning.

Ratios

Substantial research has established a relationship between smaller child-staff ratios and better social and cognitive outcomes for children.[6] Ratios were collected in different ways over the course of this study. Initially, center staff were asked to note the child-staff ratio in each classroom at four predesignated time points each day for four weeks, yielding a total of 80 time stamps for full-time programs. However, concerns about reliance on self-reported data in a high-stakes context led to our collecting ratios using sign-in/sign-out data as described below.

Background

Despite consensus about the importance of ratios as a quality indicator, little attention has been focused on how child-staff ratios can be measured most effectively. This issue of measurement effectiveness includes both efficiency and the most appropriate approaches to use in high-stakes settings. Some studies rely on self-reported questionnaire data, in which directors or staff provide ratio information (e.g., Perlman, Zellman, and Le, 2004; Rosenthal and Vandell, 1996), but the accuracy of center administrators' and caregivers' responses has not been tested. Certainly, this method permits downward reporting bias; this is a serious issue in contexts in which centers receive differential reimbursement based on a quality of care rating. Thus, initially, we focused on how to measure ratios in a way that was feasible but also reliable in a high-stakes context.

A more common approach to ratio assessment involves the use of outside observers who visit classrooms and assess child-staff ratios. Typically, a single classroom from each age group of interest in a center (e.g., infant, toddler, and preschool-age) is randomly selected. Trained

[6] This work is a summary of Le et al. (2006).

observers assess ratios in these selected classrooms several times during a short observation period (typically two to three hours in the morning) on a single day. These observations are then averaged to compute a center-level ratio for each age group (e.g., Howes, 1997; Scarr, Eisenberg, and Deater-Deckard, 1994).[7]

Although this method has the advantage of yielding objective information about centers' ratios, observations are costly, and many factors can render the observed ratios an unrepresentative sample of the ratios that children actually experience during most of their time in care. We know, for instance, that the number of children assigned to staff members is not consistent throughout the day (Kagan, 1991), so ratios measured during a given observation period may not accurately capture ratios at other hours of the day. Additionally, ratios can vary by day of the week if many parents use child-care services on a part-time basis. Finally, little is known about the extent to which ratios vary among classrooms serving children of the same age in a given center; ratio measurement in one classroom is assumed to be representative of all classrooms in that age group, but this may not be the case. The analysis described below was conducted to address four questions:

1. How strongly correlated are ratios determined through the sign-in/sign-out method with those obtained by independent observers?
2. How do ratios vary across same-age classrooms in a given center?
3. How do they vary across days?
4. How do ratios vary across hours in a given day?

Methods

Data Collection. As was the case with the analysis of the ECERS-R, we sought the largest possible data set with which to conduct these analyses. With Qualistar's help, we were able to augment the sample described in Chapter Two with additional centers that Qualistar was assessing outside of the RAND Q-QRIS evaluation. We took advantage of a little-used measurement approach to test the representativeness of ratios obtained during a typical single-classroom, two-hour data-collection period. Sign-in/sign-out data were collected in 77 centers (279 classrooms) over 10 consecutive days. Parents were asked to sign their children in and out of the child-care center as they dropped them off and picked them up. During the day, staff members were responsible for signing children and themselves in and out of classrooms. Staff members were not expected to sign in and out for brief periods of movement (e.g., restroom breaks) or for programmatic changes that involved staff and children moving as a group (e.g., class going out for a walk, into the music room). Using these data, we were able to calculate ratios for all rooms in a center on a moment-by-moment basis. These sign-in/sign-out data allowed us to test how well ratios measured in the more typical way—at selected intervals during a short time period from a single classroom on a single day—proxied these more complete data.

Observational data from one randomly selected classroom at each of three age levels in 20 of the 77 centers were collected during a two-week period prior to the collection of the sign-in/sign-out data. Observers spent 1.5 hours in each classroom and recorded the number of staff members and children present at every half-hour interval. These observational data are

[7] For example, if a center with two toddler classrooms had a child-staff ratio of 6-1 and the other a ratio of 8-1, the center could be considered in compliance with a 7-1 ratio requirement for toddler classrooms. But children in the two classrooms would be experiencing very different ratios, a point we discuss further below.

used here solely to validate the sign-in/sign-out data and are not used for any further analysis. Beyond validation, all findings reported here are from the sign-in/sign-out data.

Analytic Approach. We correlated ratio data collected through the sign-in/sign-out method with those obtained by independent observers. We then correlated ratios measured every 30 minutes from 8:30 a.m. to 10:30 a.m. in a single classroom on a single day with an overall ratio for children measured during an eight-hour timeframe in same-age classrooms across 10 days.

Findings

We first calculated center-level scores using both observer and sign-in/sign-out data and examined the relationship between ratios assessed by these different methods.

Comparing Center-Level Scores from Both Measurement Approaches. For each age group, we created a center-level score from the sign-in/sign-out data using the same computational method as employed by observers. (See Le et al., 2006—Document A.3.2 in the appendix to this chapter—for a description of how the sign-in/sign-out data were converted into ratios.) That is, we used the sign-in/sign-out data to calculate a ratio at every half-hour interval during the same timeframe and classroom as rated by observers. We then averaged the ratios at every half-hour to create a center-level score and correlated these center-level ratios with the center-level ratios obtained by averaging the measures collected by independent observers. The correlation between the sign-in/sign-out ratios and observers' ratios was .78, a strong correlation according to guidelines put forth by Cohen (1988). A t-test of the differences in ratios obtained by the sign-in/sign-out method and by observers was not statistically significant, suggesting that there were no mean-level differences in the ratios produced by the two methods. Given that both methods provide similar representations of centers' ratios, it is reasonable to assume that the ratios obtained using the sign-in/sign-out data are a good proxy for the ratios that would have been obtained using traditional observation methods. From here on, we focus only on the ratios obtained from the sign-in/sign-out data.

Assessing Child-Staff Ratios and Their Components Using Only Sign-In/Sign-Out Data. We found that the single classroom ratio underpredicts the overall average ratio. The magnitude of the difference between the two measures varies by age group, but, taking into account different ratio requirements by age group, we found that the underprediction remains fairly constant across age groups.

To determine which components—days, hours, or classrooms—contribute the most to measurement variability and therefore should be prioritized in measuring ratios, we examined the marginal distributions of ratios delineated by each of these factors. First we examined days. We found that, in general, ratios are consistent across the 10 days. Variability in mean ratios across days is fairly small, with preschool classrooms showing the largest differences (range of .95 across days) and infant classrooms showing the smallest differences (range of .53).

We found more variation in child-staff ratios when we examined them across hours. Ratios are generally similar across the same hours of different days, but they vary by hour within the same day. Ratios increase (i.e., get worse) steadily between the hours of 8:00 a.m. and 3:00 p.m., resulting in measurements that are markedly larger in the afternoon than in the morning.

We also examined the variability in ratios across classrooms in the same center. We found that, when a center has more than one classroom serving a particular age group, there are small to moderate differences in ratios between the classrooms with the fewest and most children.

For instance, for centers with two toddler classrooms, there are, on average, 1.81 additional children per staff member in the larger of the two classrooms. This difference becomes more pronounced for three-classroom centers and for the preschool-age group.

Increasing the number of days on which ratios are obtained and increasing the number of measurements in a given time period have only slight effects on improving correlations with the overall average ratio. On the other hand, increasing the time period during which ratios are collected from two hours to eight hours has a much larger effect. These results are consistent across age groups and hold for centers with one or with multiple classrooms for a given age group.

How Best to Deploy Limited Measurement Resources. Given that eight hours of ratio data collection is impractical, are ratios more accurately measured by spending more time in a single classroom or less time in more classrooms? To address this question, we examined the correlations between the overall average ratio and ratios obtained during a two-hour timeframe for one and two classrooms. Measuring ratios from an additional classroom resulted in relatively large improvements in correlations, but measuring ratios for an additional hour in a single classroom resulted in more gradual gains, suggesting that increasing the number of classrooms measured is the best way of improving ratio representativeness per unit of time.

Discussion

Our results suggest that ratios obtained during a two-hour timeframe from a single classroom on a single day are moderately correlated with ratios measured from multiple classrooms across eight hours for 10 days, although they underpredict ratios obtained using more data. The greatest source of variability compromising the representativeness of ratios measured from the single-classroom, two-hour timeframe was variability across classrooms.

That ratios in one classroom are not necessarily indicative of ratios in another classroom serving the same age group raises questions about the meaning of center-level ratios that are sometimes specified in licensing guidelines, which are obtained by averaging across classrooms serving the same age group. Differing ratios across same-age classrooms suggest that it may be advisable to also institute a "threshold" level to ensure appropriate ratios across same-age classrooms. These guidelines might specify an acceptable child-staff ratio at the classroom level and not just at the center level. This finding of variation across classrooms serving the same-age children also suggests that, in evaluating the quality of a center as a whole, a center-level average ratio may need to be supplemented with some measure of range or deviation from the average in order to provide a more complete picture of the ratios that children actually experience in their classrooms. Or, the center-level average ratio may need to depend on assessments from more than one classroom in each age range.

Our findings also have implications for improving ratio assessment. We found that ratios were best early in the day, increasing in a linear fashion to a peak between 2:00 p.m. and 3:00 p.m. across all of the age groups we studied. Variation in ratios by hour may reflect centers' attempts to concentrate personnel during certain times of the day when particular types of activities occur. For example, a center may decide that access to staff may be especially important for children during times of active learning, but less important during naptime or free play periods.

Given our findings and the possibility that centers staff more heavily during "instructional" time, the common practice of taking ratio measurements in the morning is likely to result in an underestimate of the overall average ratio; this may make sense if the purpose of

a study is to understand how ratios are related to children's cognitive outcomes. If, however, the purpose of ratio measurement is to rate centers in terms of the ratios that children actually experience over the course of a day, then it may be important to use data-collection methods that better approximate the overall average ratio.

These analyses led to changes in the way that ratios were assessed in the Q-QRIS. Consistent with the recommendations presented in this chapter, the Qualistar standard now involves eight time stamps collected by observers over the course of three days, and assessments are made in each classroom in a center.

Staff Education, Training, and Experience

The education, training, and experience of classroom staff and center directors have been linked to child-care quality and to child functioning (Howes, Whitebook, and Phillips, 1992; Arnett, 1989; de Kruif et al., 2000; Blau, 2000; Goelman et al., 2006). It is the caregivers, after all, who interact directly with children, deliver (and in many cases, plan) the curriculum, and hopefully take notice of and build on each child's interests and skills to encourage exploration of ideas, materials, and relationships.

Below, we discuss some of the key issues in defining and measuring staff education, training, and experience. Limited resources precluded in-depth analyses of these questions in the context of this study in most cases. We do present results of regression models that were intended to identify which training and education variables best predict classroom process quality as measured by the ECERS-R. We also present self-reported data on staff movement within centers. Finally, we outline future directions for research in this area.

Background

Caregiver training and education are typically assessed through three components: highest level of formal education, amount of training (which may include ECE credits taken pre- or inservice, as well as other inservice training), and years of experience working in a child-care setting.

Despite the many studies in which these indicators have been employed, there is virtually no research on how best to assess this important child-care quality component or weight and combine the different elements included in the concept. Some of the key issues in measuring staff capacity are only rarely addressed. For example, measures assume that staff remain in a classroom all day, but survey reports suggest this may not be the case (Cryer, Hurwitz, and Wolery, 2000). If staff move frequently, deciding whose background to measure is more challenging, as children may have limited exposure to any given teacher. Nor has much attention been focused on the extent to which centers rely on particular staffing models. If, for example, a center pairs teachers with B.A.s with less-well-educated ones, it may be reasonable to assume that it is the B.A.-level teacher who is overseeing curriculum and supervising the other teacher, so his or her education level sets the tone for the classroom. But little is known about staffing models; indeed, there is skepticism in some quarters that centers, facing high rates of staff turnover, can even consider such a concept. In the absence of staffing models, or in cases where two not-so-well-educated teachers staff a classroom, how should a classroom- or center-level score be derived?

Having a bachelor's degree is considered highly desirable; some researchers have found that more formal education, and a B.A. in particular, is an asset in the classroom (e.g., Burchinal et al., 2002; Howes, 1997; Tout, Zaslow, and Berry, 2005). However, a recent analysis found that a B.A. was often unrelated or related inconsistently to classroom quality and children's outcomes (Early et al., 2007).

There is also evidence that the course of study pursued toward a B.A. degree matters. For example, Dunn (1993) found that children who had a teacher with a B.A. in a child-related field, including early childhood education, outperformed children whose teacher had a B.A. in another field on the CBI Intelligence Scale. The effects of specialized training are also evident in findings that teachers with a CDA (a child development associate's degree, which is a subspecialty of an associate's degree) behave much like B.A.-level teachers on some indicators. For example, in the Florida Quality Improvement Study, both B.A.-level and CDA-level teachers demonstrated higher levels of language play and more positive management than teachers with only a high school education (Howes, 1997).

The low salaries paid to caregivers limit the number of people with a college degree who are willing to provide child care. Inservice training, that is, training provided once a person has begun to work as a caregiver, represents an attempt by many centers to compensate for lower levels of preservice education. Such training also provides all caregivers with opportunities to learn about new research and pedagogy in the field, discuss their ideas about early childhood, and learn and practice new approaches to working with young children. In the context of QRISs, attaching incentives to training is viewed as important to encouraging provider participation. Since formal education is slow and costly, QRIS designers worry that, unless training is rewarded, providers may be unwilling to be rated (see Zellman and Perlman, forthcoming, for further discussion on provider QRIS incentives).

Amount of experience is the third widely measured aspect of staff training and education. It makes sense that caregivers with more experience will be able to manage a classroom better and be better equipped to provide children with the individual attention they need (LoCasale-Crouch et al., 2007). However, the effects of caregiver experience are mixed, with some research finding that staff with more experience provide care of worse quality (Phillipsen et al., 1997).

Some limited research attention has also focused on the center director (e.g., Cost, Quality, and Child Outcomes Study Team, 1995). Her background, if it is very strong, may compensate for the lower levels of education typically found among caregiving staff. If the director is an active mentor of classroom staff, participates in the selection of curriculum materials, and conducts staff trainings, her insights and education may help to improve quality even if classroom staff themselves are not well educated. Indeed, the idea of parlaying the higher education of a few staff members into improved classroom practice is institutionalized in Department of Defense Child Development Centers, where a Training and Curriculum Specialist, who must have at least a bachelor's degree in early childhood education, works closely with classroom-based caregivers (virtually all of whom have only a high school diploma and no ECE credits when they are first hired) to improve the quality of care they provide. Burdened with neither the responsibility of running the center nor with direct care of children, the director can focus on both formal and hands-on, classroom-based training and mentoring, as well as curriculum development and materials selection and use. (See Zellman and Johansen, 1996, for further discussion of this model.) In recent years, attention has been focused on the center director's background in management and her administrative skills, on the grounds that a center must

be well run in order to provide quality care on an ongoing basis. Indeed, the National Association for the Education of Young Children includes director's management background in its assessment for accreditation. Examination of this aspect of staff background deserves more attention.

Given the general agreement on the importance of staff background as a quality indicator, surprisingly little attention has focused on how best to measure each of the components. While many research studies and quality measures rely on staff self-reports of education, these reports are flawed in several ways. While self-reported data are cost-efficient to collect, the accuracy of center administrators' and caregivers' responses has not been tested. As with self-reports of child-staff ratios, this method permits upward reporting bias—a major concern in high-stakes settings. This concern exists at the individual level as well. There is obvious reason to exaggerate one's credentials, especially when they are tied to salary and promotion. To avoid potential problems associated with relying on self-reported data, Q-QRIS T&E assessments involve extensive transcript and document reviews. This is a very time-consuming and costly process. And education is relatively easy to assess compared with inservice training. But including training in its QRIS was important to Qualistar. Focusing solely on formal education would remove incentives for staff to attend trainings and workshops and would be particularly discouraging for the many staff with low levels of formal education.

In discussions of T&E assessment, it is generally taken as a given that the staff being assessed in a given classroom spend all or nearly all their time in that classroom. However, this assumption has not been addressed in the research literature. Below, we address two questions designed to help clarify how best to collect and understand T&E data:

1. How well do the individual T&E variables relate to classroom quality? Knowing whether some are significant predictors of classroom quality while others are not would help determine whether measures of all the variables are needed and perhaps inform how they should be considered.
2. How reasonable is it to assume that staff movement is not a factor that must be considered? Linking children to staff makes sense only if their presence in a room is relatively stable.

Methods

Data Collection. Data were collected by Qualistar staff who went to the child-care sites included in the study to review and code transcripts and training certificates of all teaching staff in evaluation classrooms, as discussed in Chapter Two. Given the lack of empirical guidance on most of these questions, Qualistar simply collected extensive T&E information that could be combined in study analyses to create appropriate measures of staff training and education. Throughout our analyses, we minimized assumptions by separately examining the three T&E constructs (formal education, years of experience, and ECE credits) whenever possible. The star analyses described in Chapter Five are the major exception to this practice.

Analytic Approach. We ran regression models predicting ECERS-R score from the following five T&E variables: director's education (B.A. or not), director's administrative experience in years, the proportion of head teachers in a center who had a B.A., the number of ECE credits that head teachers had, and whether the head teacher had an ECE degree. In addition, we included an interaction term between B.A. and ECE degree in order to explore the effects of specialized child-development training above and beyond bachelor's degree attainment. A host

of covariates were included in these models to control for the backgrounds of the children cared for in these classrooms and other aspects of classroom quality. (These are listed in Table A.3.1 in the appendix to this chapter.) Wave 1 (N = 125) and Wave 2 (N = 118) data were used.

We also examined data on staff movement from the Organizational Change Survey, which center directors completed. See Chapter Two for more details including response rates.

Findings

We used classroom-level T&E data to assess the degree to which the different T&E measures relate to the ECERS-R. Although these analyses would not be definitive, they would represent a first step in examining the many different T&E variables used in this and other QRISs. Using data from the Organizational Change Survey, we address the issue of movement in a preliminary way. We also provide a brief discussion of how we entered the T&E data into the models presented in Chapters Four and Five.

Relationships of T&E Variables to Measures of Quality. If it could be demonstrated that T&E variables substitute for each other, it might be possible to collect fewer T&E variables, reducing data-collection costs. We examined which T&E variables were most predictive of quality, as measured by ECERS-R, as a first step in understanding these variables and their relationship to other measures of quality. None of the T&E variables predicted ECERS-R scores in either wave.

Staff Movement. The OCS data reveal that 82 percent of responding teachers reported that they move staff among classrooms; 67 percent reported that, in a typical week, teachers change classrooms at least once (Le et al., 2006). Since such movement is not generally considered to be good practice, there is reason to think that directors and staff would underreport rates of such movement. In these data, we also found that a number of centers employ a full-time floater teacher and some employ part-time floater teachers. Floater teachers can be assigned to any classroom; their presence during an observation did not necessarily mean that they spend substantial amounts of time in that classroom on a regular basis. The use of floater teachers indicates that some centers have created a structure that recognizes and addresses a nontrivial level of staff movement. Indeed, the floater position is designed to provide children with familiar substitute teachers. Clearly, movement of staff deserves more empirical attention. The limited data we analyzed suggest that movement of staff between classrooms is not uncommon. We will be examining our data in the future to gain more insight into staff movement. In the meantime, following past research in this area, we do not factor such movement into our models described in Chapters Four and Five.

Discussion

We need to know a good deal more about how T&E should be measured, how center directors think about T&E in staffing classrooms, and how limited resources might best be used to maximize child-care quality. The regression analyses we conducted found no links between the five T&E variables we included and the ECERS-R. Nor did they lend support to the validity of these variables as measures of quality using the limited test we applied. Unfortunately, the lack of relationships meant that the results did not provide us with guidance as to how to weight and combine the different T&E components. Analyses of the weighting and combining of T&E variables are important, as all QRISs include these variables. Testing different schemes and examining their relationships to other measures of quality will help future QRIS designers make more-informed decisions about these important matters.

In Chapter Five, we summarize cross-sectional and lagged findings about the relationships between staff and director T&E variables and other process-quality and child outcome measures. However, much work is needed to inform how T&E should be measured efficiently, reliably, and in a valid way.

Parent Involvement in Child-Care Settings[8]

Qualistar was committed to including a measure of parent involvement in its quality rating because system designers believed that quality programs should engage parents in meaningful ways and that by including the concept in its QRIS, programs would make more efforts to engage parents. This construct was the least clearly articulated of the ones initially included in the Qualistar QRIS. As we began to examine the measure that was included in the original Q-QRIS, it became apparent that much work would be needed to conceptualize appropriate parent involvement in a child-care programs and, once done, find a valid way to measure it.

In undertaking this work, we addressed the following questions:

1. How should parent involvement be measured?
2. To what extent do the parent involvement measures used by Qualistar differentiate among programs that vary in quality on other measures?
3. Does a new measure of parent involvement, the Family Partnership (FP) measure, which is based on a less traditional approach to parent involvement, relate to other measures of quality in the Q-QRIS?

Background

The notion that parents should be involved in their children's schools because they add value to both the school and their own children's education has continued to gain popularity, based on the results of a number of correlational studies that associate parent involvement with positive child outcomes in elementary-school-age children (e.g., Comer, 1984; McNeal, 1999; Stevenson and Baker, 1987; Reynolds, 1992; Muller, 1993). A number of child-care quality rating systems include parent involvement among their multicomponent indicators (see Zellman and Perlman, forthcoming, for a discussion of parent involvement measures in five pioneer QRISs).

But it is not always clear what to measure; conceptualizing and measuring parent involvement in preschool settings raises complex issues. Children who are in care for more than a few hours a week are there precisely because their parent or parents work, indicating that time available for any type of parent involvement is very limited (Shimoni, 1992). Moreover, little attention has been paid to examining how best to conceptualize, operationalize, and assess parent involvement.

In developing a measure of parent involvement to be used in high-stakes settings, several key issues must be considered. First, it is important to differentiate between provider efforts to support parent involvement and how much actual involvement occurs. This distinction is critical in high-stakes contexts so that providers that serve parents who are less likely to get involved, e.g., single parents, are not penalized for serving these populations (Kontos

[8] This work is a summary of Zellman and Perlman (2006).

and Wells, 1986; Kontos and Dunn, 1989). Identifying and measuring appropriate kinds of activities is important as well. For parents of young children, much concern focuses on basic biological functions, such as sleeping and eating. Moreover, children change very quickly in the early years. Thus, parent involvement activities often found in elementary schools, such as semi-annual parent-teacher conferences, while desirable, may not fully meet the needs of parents, caregivers, or children in child-care settings. Other approaches, discussed below, may be more appropriate.

Another measurement issue is the heterogeneous nature of child-care settings. Most child-care providers are family child-care providers that care for small numbers of children in their own homes. Compared to child-care centers, these providers tend to run informal programs; the notion of parent boards and meetings seems inappropriate to many. Parent conferences may also make less sense in a setting where parent and provider interact every day and often have close relationships. Consequently, the goals of parent involvement and its manifestations in these two quite different settings must be considered; we collected parent involvement data in homes as well as centers and report cross-type comparisons below when we have sufficient data to do so.

A key decision in measuring parent involvement is determining the appropriate respondent—provider, parent, or both. Measures that elicit responses from providers implicitly assign a substantial share of the responsibility to them for encouraging and supporting parent involvement. Staff responses may well be influenced by self-presentation biases, particularly in high-stakes settings. Therefore, attempts to collect information from parents may be warranted.

Measures that focus on parents assume that they are the key informants on issues of parent involvement, since they are the target of such efforts. If a provider takes steps to involve parents but parents are not aware of them, one could argue that the efforts are inadequate or misplaced. For example, if a provider has a policy of providing parents with community referrals and can produce a guidebook of such resources but no parents know about it, can it be said that the provider has met its obligation to provide community referrals? Some of these measures also assume that there may be more to the issue of parent involvement in preschool settings than the provision of information. Several measures (e.g., the Caregiver-Parent Partnership Scale [Ware et al., 1995] and the Parent-Caregiver Relationship Scale [Elicker et al., 1997]) focus not on the provider's parent involvement efforts *per se* but on the communication between parent and provider and on the parents' perception of the provider-parent relationship.

Our measure of Family Partnership, described below, assumes that the primary goal of parent involvement should be to help parents and providers integrate children's child-care experiences with family life. By spending even limited time in the child-care setting and through informal interaction with caregivers, parents ideally could acquire valuable qualitative information about their child's day that could help children make a smoother transition from child care to home. Parents could also learn about developmental milestones, approaches to regulation and discipline, and peer relationships, all of which could help parents better understand and relate to their child. Caregivers could learn more about the child and about the family's practices and values.

Methods

Data Collection. Data were collected from varying numbers of parent volunteers who completed one of several parent involvement measures, listed in Table 3.1. We also closely examined the ECERS-R parent involvement item; data for these analyses were already available.

Analytic Approach. We examined response patterns on the above parent instruments, focusing on percentage of responses that were not at the top of the scale (i.e., ones that were not the most positive response option possible). We compared item response patterns for parents rating centers and homes. We also correlated the parent and provider components of a new measure we devised with each of the Q-QRIS components.

Findings

When we began our analysis of the Q-QRIS, Qualistar was using the NAEYC measure as its parent involvement measure. This instrument, designed for use in the self-study process, primarily assessed parent satisfaction with their child's program. However, that measure demonstrated virtually no response variability across programs that were known to differ in quality. Subsequently, we explored a number of existing items/measures that operationalized parent involvement in different ways (see Table 3.1). Unfortunately, as summarized in Table 3.1, each measure had significant limitations, leading us to reject all of them in favor of developing our own measure, which we called the Family Partnership measure. See Zellman and Perlman (2006) for a detailed description of the analysis of the measures in Table 3.1. (This paper may be found as Document A.3.3 in the appendix to this chapter.)

The FP measure is based on the notion that productive parent involvement efforts should focus on helping parents to develop and maintain a good relationship with their child and their child's program. Child-care providers can most effectively support the first goal by helping par-

Table 3.1
Summary of Work with Existing Parent Involvement Measures

Instrument	Focus	Responses	Items	Limitations
NAEYC Parent Form (18 yes/no items used as part of NAEYC's accreditation process[a])	Satisfaction with program	292 parents	28	>90% of responses were favorable for each item
National Longitudinal Study of Youth (NLS-Y) items on parents' contacts with providers	Parent/provider contacts: phone/in-person	Approximately 214 parents (varied slightly by item)	3	Substantial differences between home and center responses
Parent-Caregiver Relationship Scale	Quality of parent-caregiver relationship	222 parents	35	Inappropriate focus on liking of and personal relationship with caregivers
ECERS-R Parent Involvement Item	Provisions for parents, e.g., involvement opportunities, sharing of child-related information, program information, parent program evaluations and decisionmaking	52 classrooms	1	Completed only by providers

NOTE: All of the analyses are based on Wave 1 data.

[a] We also tested a modified form on which the items were placed on a 5-point scale.

ents know their child better, understand how to foster their child's learning and growth, and learn more about accessible child-friendly activities. For example, a parent who is told that her child, normally a good napper, did not nap that day may be better prepared for moodiness and decide on an earlier bedtime (Owen, Ware, and Barfoot, 2000). In more elaborate interactions, a parent may begin to model the nonpunitive behavior-control approaches that well-trained caregivers display. Child-care providers can effectively support the second goal by communicating frequently with families and by providing them with opportunities to participate in many different aspects of their child's program, including but not limited to traditional forms of involvement, such as volunteering in classrooms or on field trips. The limited empirical data on this point suggest that parent-provider communication is associated with higher-quality care (e.g., Ghazvini and Readdick, 1994; Ware et al., 1995). To provide this help, the caregiver and parent have to establish a positive relationship that in some sense is a partnership formed on behalf of the child. (See Documents A.3.4 and A.3.5 in the appendix to this chapter for the FP Family Form and the FP Documentation Survey.)

An FP Family Form was completed by 1974 parents in 60 programs. The results were familiar: high levels of positive responses and little variation among providers or parents. On the 18-item form, there were only 7 items for which the level of positive response ("strongly agree" and "somewhat agree") was less than 75 percent. This positivity is especially striking because several of these items asked parents how much their own behavior had changed as a result of input from their child's provider. Positivity levels were even higher on the FP Documentation Checklist that providers completed. On only 5 of the 16 items was the level of positive response less than 90 percent; these items focused on concrete matters, such as use of family evaluations in program planning, existence of a written parent involvement plan, annual evaluation of family activities, and inclusion of family members in program planning.

The FP data reveal some noteworthy differences between home- and center-based care. In general, parents using home-based care reported even higher levels of satisfaction than parents using centers. This was particularly noticeable on items dealing with parent-caregiver interaction. Parents using home care reported feeling more comfortable and believed their ideas were better heard and heeded than those using center care.

Despite the high positivity levels, some of the variation we had sought to achieve in selecting parent items that focused on a range of parent involvement components did in fact emerge. For example, a substantial minority of parents was clearly not ready to endorse their provider's parent involvement efforts; 23 percent did not agree (strongly or somewhat) that their child's program had given them helpful ideas and information for working with their child at home; 30 percent did not strongly or somewhat agree that the program offered information about family activities and community resources.

As part of the FP development process, we examined whether early versions of the measures bore any relationship to other measures of child-care quality. In particular, we analyzed whether parent involvement, as measured by the FP, related to other Q-QRIS components and measures in our center sample. (We did not have enough homes to conduct analyses on them.) We examined the correlations between the FP and the learning environment (measured by the ECERS-R or ITERS), teacher credentials (measured by years of teaching experience, education level, and ECE credits), and child-staff ratio. Points were awarded for each item on the parent

survey and provider survey. Totals were then calculated for the parent survey, provider survey, and overall score. Parent survey points were averaged to create a provider-level score. We correlated total points, parent survey points, and provider survey points with the other indicators in turn. In general, all three FP measures were strongly related to the other quality measures.[9]

These findings underlined the potential of the FP to be viewed and treated as an indicator of quality. Despite strong positivity biases, parents appear to be discriminating in their responses, awarding higher ratings to programs that are measurably better based on other indicators. Even the more-positive program responses to the provider survey appear to discriminate quality. Of course, the correlational nature of these data does not clarify whether parent involvement improves child-care quality, is a product of other aspects of quality, or simply co-exists with other quality components. It is also worth noting that the correlations between the quality indicators were stronger for parent reports of parent involvement than they were for information collected from child-care providers. Several factors might explain this pattern, including fewer items on the provider survey that would artificially depress strength of correlations in comparison to the parent survey, more motivation among providers than among parents to present the program positively in a high-stakes setting, and perhaps, among the lowest-quality providers, a lack of understanding of the nature of good relationships with parents.

Clearly, more work needs to be done with the FP measure, particularly if it is to be adopted in high-stakes settings. In particular, the relationship of this new measure to other measures of related concepts, such as parent involvement/parent engagement, needs to be explored. But the measure seemed promising enough to include in our QRIS analyses, described in Chapter Five.

Discussion

The long process to develop the FP meant that parents responded to a different parent involvement measure each time they were asked. Analyses of this component described in Chapter Five are therefore limited. The new FP measure is promising in that experts at Qualistar believe it captures key parent involvement components that make more sense in child-care settings than the more traditional measures of parents showing up for meetings or performances. It also captures some variability in parent responses. The measure is also valuable because it allows some flexibility to programs (e.g., a "welcoming environment" might include a variety of features) and also appears to work in different settings (e.g., centers and homes). The data presented here provide preliminary support for the validity of the measure, although different patterns of relationships between versions indicate more work is needed. In particular, there needs to be an examination of the content of the FP by a broader panel of experts, and an analysis of the relationship of the FP measure to other accepted measures of parent involvement.

[9] These correlations were based on smaller numbers and unimputed data, so they are not presented here. Correlations between the FP measures and other Q-QRIS measures using the full sample and imputed data may be found in Tables 4.2 and 4.3

Accreditation

Accreditation in the Q-QRIS relies on another quality rating scheme, usually the National Association for the Education of Young Children's accreditation process.[10] That process is time-consuming and complex, as it involves staff, administrators, and parents in a self-evaluation process. One of our goals was to assess the contribution of accreditation to the Q-QRIS. Another was to determine the contribution of accreditation status to classroom quality and child outcomes.

Background

The NAEYC accreditation process was first established in 1985, at a time when there was little federal leadership in the development or support of early child-care and education programs.[11] In the 1980s, quality was something that was discussed but rarely measured. Caregivers usually depended on inservice training for their knowledge and skills. Accreditation was viewed as a tool to upgrade the way that quality in all aspects of care was understood. By setting standards of excellence based on a set of consensually derived accreditation criteria, the field would have a target of excellence and a way of recognizing those who attained it. Moreover, by designing a participatory system to achieve accreditation, the process could help caregivers become more actively involved in thinking about quality, which hopefully would lead them to take steps to improve it in their daily work. The required self-study process would involve caregivers, administrators, and parents in assessing the center and working to improve its quality.

The accreditation process involves four steps: (1) enrollment in self-study, (2) application *after* self-study, (3) candidacy (which involves preparation for a site visit after NAEYC screening of the self-study materials), and (4) meeting and maintaining standards (this step includes the site visit).

It is based on 10 standards, including relationships, curriculum, teaching, assessment of child progress, health, teachers, families, community, physical environment, and leadership and management.[12]

Successful accreditation requires that programs meet all required criteria and meet all standards at 80 percent, and that each classroom meet 70 percent of all criteria. Some critics are uncomfortable with these percentages: They argue that standards are not rigorous in practice (see Zellman and Perlman, forthcoming, for state views on accreditation in their QRISs).

Noteworthy as well is the absence of any mention of child-staff ratios or group sizes in the accreditation standards, considered by many to be critical to assessments of quality. While accreditation is separate from the ECERS-R, some of the key ECERS-R constructs (e.g., interaction, space, and furnishings; personal care routines; parents and staff) are assessed as part of the standards. There are different emphases, however. For example, the accreditation standards

[10] While Qualistar accepts accreditation by other agencies as discussed above, most accredited programs are accredited by NAEYC. For ease of presentation, we use NAEYC accreditation to describe accreditation in this chapter.

[11] NAEYC introduced a revised accreditation process in September 2006. The revised process is based on 10 new Early Childhood Program Standards that programs must meet. The process has also changed: There are now four steps and new forms, terminology, and deadlines. The process still requires annual reports, involves random unannounced visits, and requires programs to report major status and program changes. See NAEYC (no date).

[12] See NAEYC (no date) for further detail.

focus on teacher qualifications, whereas ECERS-R focuses on staff supervision and staff personal and professional needs.

Empirical data that support the validity of accreditation data as an indicator of quality are limited. Zan (2005) evaluated 116 NAEYC-accredited preschool classrooms. While the mean total ECERS-R score was 5.77, classrooms exhibited a wide range of scores from inadequate to excellent on 11 curriculum-related items. Thirty-four classrooms scored in the inadequate or minimal range (1–3) on at least one of the curriculum-related items, and 18 scored in this range on four or more of these items. Zan argues that the results demonstrate that the previous NAEYC accreditation criteria did not adequately address the quality of the curriculum. Whitebook (1996) reviewed a number of studies that examined accredited centers on other quality indicators. She found that quality levels were variable; many accredited programs are only adequate. She concluded that NAEYC accreditation "is not a firm criterion of excellence" (p. 43). Gerber (2003) examined predictors of sensitive caregiving in child-care centers. Study results revealed that among a range of teacher and setting variables, accreditation status emerged as the most significant predictor of both attuned and harsh teacher behaviors. Gerber notes that these findings highlight the fact that accredited centers have more resources that contribute to more-sensitive caregiving.

The realities of child-care underfinancing limit the number of accredited centers; this reality may dilute the value of accreditation. Programs that decide to apply for accreditation, and particularly those that succeed, are usually better endowed than most, charging higher fees and paying higher wages (Whitebook, 1996). Indeed, some have suggested that accreditation does little more than further distinguish programs that are good already. The fact that well under 10 percent of civilian child-care programs are accredited lends credence to this argument.[13]

In developing its Q-QRIS, Qualistar wanted to recognize that accreditation may signal a particular commitment to high-quality care and considerable dedication and effort to undertake and complete the process. It sought to reward programs that have undergone the self-study process, which brings staff together to pay attention to and improve quality and to provide incentives to others to begin that process. In addition, there was discussion among Qualistar planners that at some point in the future the Q-QRIS and NAEYC accreditation could become components of the same system; achieving accreditation might translate directly into a high rating on Q-QRIS or vice versa.

At the same time, there were concerns at Qualistar that including accreditation as a Q-QRIS component would add substantial burden and cost to its rating system without providing substantial advantage. A particular concern early on was that the Q-QRIS would be reliant on the processes of another organization. Another concern was the long wait (at that time) before a center could have a validation visit. Some at Qualistar were uncomfortable with a scenario in which all other Q-QRIS components were rated but a provider had to wait for a final Q-QRIS point rating until the NAEYC validation visit had been completed and the accreditation decision had been made. A number of people at Qualistar argued that if a center or family child-care provider achieved high levels on the other four Q-QRIS components, accreditation would be redundant. Moreover, if accreditation were given substantial weight in the overall Q-QRIS score, it would penalize providers that lacked the necessary funds to

[13] Many accredited programs are part-day pre-kindergarten, Head Start, and U.S. Department of Defense programs.

undertake the accreditation process. Some argued that the time and resources that would go to seeking accreditation might be better spent on improving the other dimensions of quality assessed in the Q-QRIS.

Qualistar reconciled these different viewpoints by including accreditation in the Q-QRIS, but assigning it relatively little weight in the total Q-QRIS score. Our analyses were designed to better understand the role that accreditation plays in the Q-QRIS. With some data on its functioning relative to the other components, Qualistar would be in a better position to decide whether to continue to include accreditation or drop it from the Q-QRIS.

We conducted analyses that addressed four key questions:

1. Did the percentage of accredited centers increase among participating providers over the course of the Qualistar study?
2. Is there a relationship between accreditation status and overall star rating on the Q-QRIS?
3. What is the relationship between accreditation status and other indicators of quality, including the other Q-QRIS components and the process measures that we used to validate the Q-QRIS?
4. Is there any relationship between a center's status (profit versus nonprofit) and its accreditation status?

Methods

Data Collection. The full sample of child-care centers available from the Qualistar evaluation was used for these analyses.

Analytic Approach. To conduct these analyses, we compared accredited and nonaccredited centers on their overall star rating. We conducted t tests of differences in mean scale and subscale scores on the ECERS-R, the training and education items, ratios, the Family Partnership measure, the Caregiver Interaction Scale (Arnett, 1989), and the Pre-Kindergarten Snapshot (Howes, 1997) for accredited and nonaccredited centers. We also examined accreditation status by profit status.

Findings

Our analyses of the effect of accreditation were severely hampered by the small number of centers in the study sample that were accredited. Of the 65 Wave 1 centers, just 7 (11 percent) were accredited, as shown in Table 3.2.

Not surprisingly, given the resources and support provided to participating centers through the intervention, the number of accredited centers in our sample increased from 7 in Wave 1 to 11 in Wave 2. These 11 accredited centers represent 21 percent of the sample of 52 centers. By Wave 3, no additional centers in the sample were accredited.

At Wave 1, none of the 30 lowest-quality centers (those with star ratings of 0 or 1 out of 4 stars) were accredited. Three of 20 centers with a 2-star rating were accredited; four of the 14 3-star centers were. Only 1 center received 4 stars; it was not accredited. Using Fisher's exact test, this pattern of results did not attain statistical significance at the p. <. 025 level, suggesting that there is no relationship between accreditation status and star level.

At Wave 2, none of the 10 centers with star ratings of 0 or 1 were accredited. Three of the 14 2-star centers (21 percent) were accredited. Six of the 24 3-star centers (25 percent) were accredited, and both centers with a 4-star rating had achieved accreditation. The 11 accredited

Table 3.2
Accreditation, by Star Rating, Wave 1–Wave 3

	Star Rating						
	0	1	2	3	4	Total	p. value
Wave 1							
Accredited	0	0	3	4	0	7	
Nonaccredited	20	10	17	10	1	58	
Total	20	10	20	14	1	65	<.06
Wave 2							
Accredited	0	0	3	6	2	11	
Nonaccredited	5	7	11	18	0	41	
Total	5	7	14	24	2	52	<.06
Wave 3							
Accredited	0	0	0	8	3	11	
Nonaccredited	2	5	10	20	1	38	
Total	2	5	10	28	4	49	<.025

centers all had 2, 3, or 4 stars. None of the nonaccredited centers received 4 stars, and none of the accredited centers had fewer than 2 stars. Again, using Fisher's exact test, this pattern of results did not attain statistical significance at p. < .025.

At Wave 3, the 11 accredited centers all had 3 or 4 stars; just 1 of the nonaccredited centers had 4 stars. This pattern of results attained statistical significance (p. < .025) using Fisher's exact test. See Table 3.2 for these data.

Given Qualistar's interest in the value of including accreditation in its Q-QRIS, we examined the relationship between accreditation status and other indicators of quality, including the other Q-QRIS components and the process measures that we used to validate the Q-QRIS. In the center sample for each wave, we examined the relationship between accreditation status and other indicators of quality by testing differences in means on each variable by accreditation status. Mean differences between accredited and nonaccredited centers on the 14 variables, shown in Table 3.3, were compared.

T-tests of differences in means by accreditation status for the Wave 1 data revealed that none of the 14 means were significantly different at the .025 level. T-tests of differences in means by accreditation status revealed significantly higher ECERS-R scores in accredited centers in both Waves 2 and 3. See Table 3.4 for these data.

On the FP, we found no significant differences in mean provider survey points between accredited and nonaccredited centers. However, we did find significant differences in parent survey points favoring accredited centers, as shown in Table 3.5. As part of the NAEYC accreditation process, centers must survey parents about parent involvement (recall from the previous discussion that Qualistar initially measured parent involvement using NAEYC's measure). Going through the accreditation process may put mechanisms in place that raise providers' FP scores and that are reflected in the findings.

Finally, we examined whether for-profit or nonprofit centers were significantly more likely to be accredited. Findings across waves revealed that profit status was unrelated to accreditation.

Table 3.3
Indicators of Quality Examined

QRIS Components
Classroom Environment
ECERS Wave 1 provider mean
Staff Training and Education
Number of ECE credits—head teacher
Years of teaching experience—head teacher
Number of ECE credits—director
Years of administration experience—director
Average Class Ratio
Process Measures
Caregiver Interaction Scale
Detachment
Permissiveness
Positive relationship
Punitiveness
Pre-Kindergarten Snapshot
Class mean for proportion of observations in which there was responsive adult involvement
Class mean for the quality of peer play
Class mean for nonplay
Class mean for cognitive play

Table 3.4
Relationship Between Accreditation Status and ECERS-R Score, by Wave

	Wave 1 (N =65)				Wave 2 (N = 52)				Wave 3 (N = 49)			
Accreditation Status	n	Mean (SD)	t	p.	n	Mean (SD)	t	p.	n	Mean (SD)	t	p.
Nonaccredited	58	4.21 (1.25)	–2.25	.028	41	4.91 (1.02)	–2.51	.020	38	5.21 (.73)	–3.45	<.001
Accredited	7	5.30 (.83)			11	5.71 (.60)			11	5.96 (.42)		

NOTE: SD = standard deviation.

Table 3.5
Relationship Between Accreditation Status and FP Scores

	Accredited Centers (N = 11)		Nonaccredited Centers (N = 29)			
FP Measure	Mean	SD	Mean	SD	t	p.
Parent survey	21.55	2.38	17.17	7.10	–2.91	<.006
Provider survey	23.91	1.22	22.62	3.58	1.70	<.10

NOTE: FP Version 5 was used in these analyses.

Accreditation was included as a predictor in the regression models described in Chapter Four. These models include all of the Q-QRIS components as well as a number of covariates (e.g., family income) necessary to understand the link between child-care quality and child functioning. Both cross-sectionally (at Wave 2) and across time (from Wave 1 to Wave 2 and from Wave 2 to Wave 3), accreditation status did not predict any of the child outcomes examined in this study. This is not surprising given the low numbers of accredited centers in the study sample and the general lack of association between the Q-QRIS components and either the process variables or child outcomes.

Discussion

Our data indicate that accreditation status tends not to be associated with the Q-QRIS star rating. Only in the smallest wave (Wave 3) did we find a significant association between star rating and accreditation status. Across three waves of data and many variables, we found a few significant associations between accreditation status and other indicators of quality, although the differences are inconsistent across waves. Accreditation status is not related to any of the process variables.

Given the cost of accreditation and the dependence it requires on another quite time-consuming quality rating scheme, these findings suggest that Qualistar may want to consider dropping accreditation from its Q-QRIS. Currently, it counts for only 20 percent of what each of the other four components contributes to Qualistar's provider rating. This diminished value relative to the other four Q-QRIS components is Qualistar's way to encourage accreditation while also recognizing that some centers cannot afford the time or cost of undergoing the self-study process. These centers may, in fact, benefit more from focusing limited time and resources on other quality-improvement efforts; research is needed to test this notion. However, our ability to draw conclusions about the contribution of accreditation to the Q-QRIS is limited by the very small sample of accredited centers in this sample.

Conclusions

When RAND began working with Qualistar Early Learning, it became apparent that the Q-QRIS components needed to be examined individually before we could sensibly examine the Q-QRIS as a whole. Because of problems we identified at the component level, changes were made to the Q-QRIS. This resulted in improved measures of the individual components but compromised our ability to conduct longitudinal analyses of the Q-QRIS as a whole.

This chapter describes some of our analyses of the components. The measure of parent involvement and ratios received the greatest amount of attention. Our work has affected the measurement of both Q-QRIS components. However, as discussed throughout this chapter, more research is clearly needed on how to operationalize and measure several of the components. With regard to the ECERS-R/ITERS/FDCRS, fundamental questions remain as to whether they are the optimal measures of the classroom environment. Future research should examine which indicators of training and education should be measured and how components should be combined across staff within a room. Parent involvement should be further analyzed at both the conceptual and measurement levels. Given the small number of accredited centers available in this sample, further research on the value of incorporating accreditation into the Q-QRIS and into QRISs in general is also needed.

CHAPTER FOUR

Relationships Among Components and Component Changes over Time for Center Providers

This is the first of two chapters that examine the Q-QRIS components simultaneously. We first examine the relationships among the Q-QRIS components, looking at classroom-level, provider-level, and a mix of classroom-level and provider-level correlations for the center sample. Then, we examine improvement over time in the quality of participating providers as measured by key Q-QRIS components.

Each of these sets of analyses addresses the validity of the Q-QRIS. Analyses of the interrelationships among the component measures of the Q-QRIS speak to its internal structure. Ideally, we would find moderate correlations, which would indicate relatedness but not redundancy. Since providers volunteered to participate in the study and received quality-improvement support, it seems reasonable to assume that their quality might improve over time. If the Q-QRIS is measuring quality, we would therefore expect improved scores on the Q-QRIS components. Ideally, we would have examined changes in the criterion measures in the same way. However, since these were taken in only one classroom and the classrooms varied by wave, we could not conduct a similar analysis (i.e., changes across waves could be due to the selection of different classrooms). Examining change over time and the relationship of any change to the Q-QRIS components would support the Q-QRIS as a measure of quality.

The two sets of analyses reported in this chapter address the following questions:

1. How closely do the Q-QRIS indicators relate to each other?
2. Did the quality of participating providers that remained in the study from Wave 1 through Wave 3 improve?

Methods

We first correlated the Q-QRIS measures. Because these measures represent different levels of provider functioning (classroom- and center-level), we conducted two sets of correlations for each wave, first at the classroom level and then at the provider level. A third set of correlations included both classroom- and provider-level variables.

We then examined changes over waves in mean Q-QRIS measures for providers that remained over the course of the study. We examined the significance of differences in mean scores over waves using repeated measure analyses.

Correlations Among the Q-QRIS Components

We examined the relationships among the Q-QRIS measures at the classroom level by correlating the average classroom ratio, classroom ECERS-R scores, and teacher credentials (measured by whether the head teacher had a B.A., head teacher ECE credits, and head teacher years of teaching experience). At the provider level, we correlated parent and provider FP survey scores, whether the director had a B.A., the director's administrative experience, and whether the center was accredited.

Finally, we correlated provider-level variables—FP parent points, FP provider points, whether the director had a B.A., director administrative experience, and provider accreditation status—with classroom-level variables—average classroom ratio, average classroom ECERS-R score, whether the head teacher had a B.A., head teacher ECE credits, and head teacher teaching experience.

As shown in Tables 4.1, 4.2, and 4.3, the measures correlated moderately well at Wave 1, particularly at the classroom, mixed classroom, and provider levels, where the sample sizes are larger. Ratios and ECERS-R scores were negatively related, as would be expected. Head teacher ECE credentials and teaching experience also correlated with ECERS-R scores. FP parent and provider points were significantly and positively associated with director education level (as shown in Table 4.2). As shown in Table 4.3, both are negatively associated with average ratios, as expected, although only the correlation for the parent survey is significant. Both FP measures correlate positively with ECERS-R scores. The parent survey measure is also positively associated with head teacher ECE credits. Director education and experience are positively related to ECERS-R scores and to teacher ECE credits; director education level is also positively related to head teacher years of experience. Finally, accreditation status is significantly related to ECERS-R scores. Results for Waves 2 and 3 generally show similar patterns, although we found fewer significant correlations, which may be due in part to smaller sample sizes. (Tables of correlations for Waves 2 and 3 may be found in Tables A.4.1–A.4.6 in the appendix to this chapter.)

These correlations suggest that the Q-QRIS components are measuring related but different concepts and that they relate to each other in expected ways.

Table 4.1
Wave 1 Correlations Among Classroom-Level Q-QRIS Components

Q-QRIS Component	Average Class Ratio	Mean ECERS-R	Head Teacher Education Level	Head Teacher ECE Credits	Head Teacher Years of Experience
Average class ratio		−0.25* (n=148)	−0.14 (n=148)	−0.13 (n=148)	0.001 (n=148)
Mean ECERS-R			0.10 (n=148)	0.35* (n=148)	0.28* (n=148)
Head teacher education level				0.30* (n=148)	0.05 (n=148)
Head teacher ECE credits					0.29* (n=148)
Head teacher years of experience					

* Significance = p. < .025.

Table 4.2
Wave 1 Correlations Among Provider-Level Q-QRIS Components

Q-QRIS Component	FP Parent Survey Points	FP Provider Survey Points	Director Education Level	Director Administrative Experience	Accreditation
FP parent survey points		0.69* (n=50)	0.55* (n=50)	0.18 (n=50)	0.23 (n=50)
FP provider survey points			0.41* (n=50)	0.13 (n=50)	0.23 (n=50)
Director education level				0.41* (n=50)	0.10 (n=50)
Director administrative experience					0.08 (n=50)
Accreditation					

* Significance = p. < .025.

Table 4.3
Wave 1 Correlations Among Mixed Classroom- and Provider-Level Q-QRIS Components

Q-QRIS Component	Average Class Ratio	Mean ECERS-R	Head Teacher Education Level	Head Teacher ECE Credits	Head Teacher Years of Experience
FP parent survey points	–0.39* (n=125)	0.70* (n=125)	0.13 (n=125)	0.38* (n=125)	0.13 (n=125)
FP provider survey points	–0.27 (n=125)	0.55* (n=125)	0.01 (n=125)	0.29 (n=125)	0.11 (n=125)
Director education level	–0.18 (n=148)	0.49* (n=148)	0.09 (n=148)	0.32* (n=148)	0.25* (n=148)
Director administrative experience	–0.08 (n=148)	0.36* (n=148)	0.13 (n=148)	0.25* (n=148)	0.21 (n=148)
Accreditation	–0.16 (n=148)	0.34* (n=148)	–0.07 (n=148)	–0.21 (n=148)	–0.03 (n=148)

* Significance = p. < .025.

Improvement Over Time in Center Quality

As discussed in Chapter One, the QRIS logic model posits improvements over time in participating providers. This hypothesis seemed reasonable for Q-QRIS providers on two counts. First, providers volunteered to participate in the study, at least in part, to benefit from Qualistar's quality-improvement efforts; these are providers that wanted to improve. These QI efforts involved extensive feedback about each provider's performance and suggestions for improvement. Second, as part of HB 1297 and other ongoing interventions, participating providers received approximately $3,500 per classroom.[1] Thus, these are providers that wanted to improve and that received support for doing so. However, any improvements in quality that we

[1] In the original study design, the amount of resources going to providers that were randomly assigned to the treatment condition was substantially higher. Centers that serve high-socioeconomic-status families and programs randomly assigned to the control group in the original study design received much less support (approximately $750 per classroom) for improvement efforts

observed could not be assumed to be solely a product of the intervention, since the intervention was not part of a randomized experiment.

Improved quality over the course of the study would enable us to examine the extent to which the Q-QRIS components as well as the process-quality indicators are associated with any improvements in provider quality over time. Findings that improvements on the Q-QRIS components at one time are associated with improvements in child outcomes at a later time would lend support to the validity of the Q-QRIS. They would also provide qualified support for the effectiveness of the QI interventions carried out by Qualistar.

The data on mean scores on Q-QRIS indicators across waves in Table 4.4 are based only on the centers for which we had data across all three waves of the study.[2] We found few significant improvements in quality indicators over time. Significant effects were almost exclusively limited to ECERS-R scores. ECERS-R scores increased significantly from those at Wave 1 at both Wave 2 and Wave 3, improving by almost one full point. Variation in these scores also decreased substantially. ECE credits increased slightly for head teachers in centers from Wave 1 to Wave 3 but declined in Wave 2. Years of teaching experience increased for center-based head teachers, although the increase was very modest. Average ratios decreased steadily, although the decreases, not surprisingly, were not large, since ratio declines represent one of the most costly quality improvements. Moreover, the providers met licensing ratio standards to begin with and, therefore, had little incentive to make such costly quality improvements.

The findings of significant improvements in ECERS-R scores in the absence of significant improvements in the other structural quality indicators shown in Table 4.4 may indicate that some quality improvement can occur without expensive structural changes, although there may be a ceiling on the amount of such change. It would be useful to examine ECERS-R

Table 4.4
Changes Over Waves in Key Quality Indicators for the 43 Centers That Remained in the Study for All Three Waves

	Wave 1				Wave 2				Wave 3			
Indicator	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
ECERS-R class mean	4.63	1.24	2.14	6.65	5.16*	.94	2.67	6.79	5.53*	.59	3.78	6.37
Head teacher ECE credits	13.89	12.07	0	41.72	11.63	9.32	0	43.00	15.88	12.9	0	57.50
Head teacher teaching experience	8.47	5.60	0	22.50	8.62	4.91	1.15	26.57	8.79	5.47	2.00	26.36
Average class child-staff ratio	6.07	1.69	3.41	11.19	5.82	1.43	3.01	9.50	5.77	1.31	3.27	9.62

NOTES: The values reported above are based on imputed data. Min = minimum range; Max = maximum range. "Head teacher" does not necessarily refer to only one staff member in a given classroom. Head teacher refers to the primary caregiver(s) in a classroom, and there may be several. For instance, approximately 38 percent of the Wave 1 sample had more than one head teacher in a class. When a classroom had more than one head teacher, the education variables (e.g., ECE credits, teaching experience, A.A. degree, and B.A. degree) were averaged to produce one class score.

* Significance = p. < .025.

[2] Six centers had Wave 1 and Wave 3 data only. Compared with the other "remaining" centers, these centers had significantly lower mean class ECERS-R scores (p. < .01), were significantly less likely to have teachers with B.A. degrees (p. < .01), and were significantly less likely to be accredited (p. < .01).

scores over time for those providers that improved, in order to identify the subscales on which change occurred. It would also be useful to examine the characteristics of centers and classrooms that improved. This might shed some light on the change process and the degree to which Qualistar's quality-improvement efforts account for improvements on these measures. Such analyses are beyond the scope of this study.

These findings of some improvement over the course of the study are consistent with an intervention of unknown magnitude and varying objectives provided to centers and homes that volunteered to participate in a study designed to improve quality.

Conclusions

The correlational data indicate that the Q-QRIS components are moderately related and that these relationships are in the expected directions. These findings lend support to the validity of the Q-QRIS as measures of quality.

The change analyses indicate that those providers that remained in the study through Wave 3 improved their quality, although improvements were largely limited to ECERS-R scores. Nevertheless, they do suggest that providers changed—for the better—over the course of the study. These results enable us to examine, in Chapter Five, whether these improvements are reflected in the process and child outcome variables.

CHAPTER FIVE

Relationships of Q-QRIS Components and Star Ratings to Process-Quality and Child Outcome Measures for Center Providers

In this chapter, we focus on three additional aspects of the validation process: the relationship of the star ratings and the individual Q-QRIS components to the measures of process quality we employed, the relationship of the star ratings and the individual Q-QRIS components to child outcomes, and the relationship of the process-quality measures to child outcomes. As part of these analyses, we examine whether these relationships vary for particular subgroups of children. In these analyses, we examine whether there were improvements over time in child outcomes and, if so, whether these improvements are associated with improved quality ratings as measured by the Q-QRIS components and the star ratings in both cross-sectional and lagged analyses. We also examine the extent to which the star ratings, the individual components, and the process-quality measures relate in substantially similar ways to the child outcomes.

The analyses address six questions:

1. What are the relationships between the Q-QRIS component measures and the measures of process quality?
2. What are the relationships between the star ratings and the process-quality measures?
3. What are the relationships between the Q-QRIS component measures and measures of child outcomes?
4. How well do star ratings predict child outcomes?
5. To what extent are the relationships between the component measures and child outcomes similar to those between the process-quality measures and child outcomes?
6. Do the relationships between the component measures and the process and outcome measures vary for particular subgroups of children?

We begin the chapter with an analysis of attrition among centers that joined the study in Wave 1, followed by a presentation of statistics for variables included in the models of Wave 1. We examine the relationship of the Q-QRIS components to the process-quality measures. Then we look at the relationship of the components and the process-quality measures to child outcomes. The final sets of analyses examine the relationship of the star ratings both to child outcomes and to the process-quality measures. We end the chapter with a discussion of factors that may be contributing to the limited effects we found in our data.

Methods

To examine how child outcomes relate to process quality, star ratings, and the individual Q-QRIS components, we used linear regression analysis. We fit these models using individual child data, with all the children from the same classroom receiving the same values on each of the scales, and we used an adjusted standard error estimate to account for possible correlation among responses from children within the same provider. We also standardized child outcome scores and provider-quality indicators so that the reported coefficient is the expected difference in standard deviation units (SDs) in the outcome measure for a one-SD increase in the quality components.[1]

We examined the relationships both contemporaneously and over time. The first set of analyses, which we refer to as cross-sectional analyses, regressed child outcomes on quality indicators while adjusting for child- or provider-level background characteristics (to be described shortly). We conducted this analysis for each wave in order to understand the robustness of results.

The second set of analyses, which we refer to as lagged analyses, examined how Q-QRIS components at a previous wave are related to outcomes in a subsequent wave. The rationale for such an analysis stems from the hypothesis that aspects of provider quality, measured by the star ratings and by the individual Q-QRIS components, may have "delayed" effects on child functioning, such that it takes time before the full effects of provider quality on outcomes are manifested.

The regression models included quality indicators that were part of the Q-QRIS system or process-quality measures. The process-quality measures from the CIS included scales relating to the positiveness, punitiveness, detachment, or harshness of the observed teacher. For the Pre-K Snapshot measure, the scales included the extent to which observed children engaged in peer play, cognitive play, nonplay, and play in which there was responsive adult involvement.

The regression models also included several variables that were likely to be related to children's social and cognitive functioning as covariates. These variables included the child's age at assessment, family income, hours per week the child attends the provider, the length of time the child has been attending the provider, whether a parent has a B.A., the minority status of the child, whether the child's family speaks another language besides English, whether the child has learning problems, the child's gender, and the parents' child-rearing style.[2] In addition, we included provider-level covariates, including whether the program is a Head Start program, whether the provider is a nonprofit organization, and the level of intervention intensity as determined by Qualistar Early Learning.[3] For the lagged analyses, we used a covariate-

[1] In Wave 2 and Wave 3, two different versions of the FP measure were administered (versions 4 and 5). Because these versions were on different metrics, we standardized the FP scores within version within wave for our analysis.

[2] The income variable was originally on a 7-point scale, where 1 represented "$1–$5,000," 4 represented "$21,001–$50,000," and 7 represented "more than $250,000." We converted these variables from fixed categories to annual income using the midpoint of the heading descriptors as a guide. For instance, a parent response of "4" for a particular income category was converted to $35,500. Parenting style was captured by the Block Child-Rearing Practices Report (Block, 1965, 1980). Two subscales, nurturance and restrictiveness, are included. (See Document A.2.12 in the appendix to Chapter Two for these items.)

[3] According to Qualistar Early Learning, the definition for each intervention level is as follows: Low intervention meant that the site either received no intervention or received only funds but no technical assistance. These were providers that were usual-care sites in Wave 1 or were sites that received only incentive money to participate. Medium intervention referred

adjustment-regression approach, in which the child outcome score from the previous wave was included in the model. This allowed us to understand how provider quality may be related to gains in social and cognitive functioning.

When sample sizes were sufficient, we also conducted a series of sensitivity analyses to explore whether relationships between child-care quality and child functioning varied for different subgroups of children. Specifically, in addition to analyses based on the full sample, we conducted parallel analyses using three different subsamples: children whose family income fell below the median for this sample ($37,500), children whose family income fell below the 25th percentile for this sample ($17,500), and children who had greater than median exposure to their child-care center based on how long they had attended and how many hours per week they spend at their center (i.e., at least 32 hours of attendance and at least 12 months in the program). This allowed us to explore whether the outcomes of children who were economically disadvantaged or who had received sustained exposure to the improvement efforts demonstrated greater benefits.

While imputation allowed us a larger sample size in some cases, the problem of small sample sizes remained when the analyses focused on relationships involving family child-care providers or on relationships between process quality and Q-QRIS components. Because the sample sizes for the home analyses lacked the power to reliably estimate coefficients, we generated correlations rather than regression analyses, which we describe in Chapter Six.

Findings

Attrition Analysis

Sixty-five centers joined the study at its inception; by the end of the study, 49 (75 percent) remained. Seven providers remained in the study for just one wave; nine remained for two waves.

We examined whether and how the centers that left the study differed from those that stayed the course. Table 5.1 presents descriptive statistics for Wave 1 data from centers that remained in the study for one, two, or all three waves.

Table 5.2 compares centers that left the study after one wave with centers that stayed on longer on key Wave 1 Q-QRIS measures.

The data reveal that on virtually all of the Q-QRIS indicators at Wave 1, the centers that left the study after Wave 1 differed significantly from those centers that remained in the study for more than one wave. In every case, centers that dropped out after Wave 1 were lower in quality, had lower ECERS-R scores, had head teachers with fewer ECE credits and fewer years of teaching experience, and had directors who were less likely to have a B.A. degree.

Table 5.3 compares quality indicators for centers that remained in the study for two waves against those that remained for all three.

In some contrast to the Wave 1 dropouts, once the comparison focuses on Wave 2 characteristics for the centers that remained past Wave 1, there was just one significant difference

---

to 1,297 sites that received an average of $3,500 in training and assistance resources per class each year. High-intervention sites received coaching, capital improvement grants, differential reimbursement, and $7,500 in training and assistance funding per class each year.

Table 5.1
Descriptive Statistics, by Wave Present

	Present During 1 Wave (N = 7)		Present During 2 Waves (N = 9)		Present During 3 Waves (N = 49)	
Variable	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Mean ECERS-R score	2.97	.54	4.68	.82	4.45	1.29
Number of head teacher ECE credits	2.27	4.40	13.65	7.22	14.55	12.18
Percentage of head teachers with a B.A. degree	0	0	.21	.30	.11	.24
Head teacher years of teaching experience	3.69	2.00	9.61	5.09	8.23	5.35
Years of director administrative experience	3.71	4.11	6.27	6.07	7.01	5.54
Percentage of directors with a B.A. degree	0	0	.42	.39	.49	.51
Average class ratio	7.04	2.93	5.97	1.54	6.11	1.67
Percentage accredited	0	0	0	0	.14	.35

NOTE: The analyses presented above are based on Wave 1 data.

Table 5.2
Comparison of Centers That Dropped Out After One Wave with Those That Stayed Longer

	Dropped Out After Wave 1 (N = 7)		Stayed After Wave 1 (N = 58)		Comparison	
Variable	**Mean**	**SD**	**Mean**	**SD**	**t value**	**p. value**
Mean ECERS-R score	2.97	.54	4.49	1.22	3.24	<.001
Number of head teacher ECE credits	2.27	4.40	14.41	11.50	5.41	<.001
Percentage of head teachers with a B.A. degree	0	0	.13	.25	3.93	<.001
Head teacher years of teaching experience	3.69	2.00	8.45	5.29	4.63	<.001
Years of director administrative experience	3.71	4.11	6.89	5.58	1.46	NS
Percentage of directors with a B.A. degree	0	0	.48	.48	7.62	<.001
Average class ratio	7.04	2.93	6.09	1.64	–.84	NS
Percentage accredited	0	0	.12	.33	2.80	<.001

NOTES: The analyses presented above are based on Wave 1 data. SD = standard deviation; NS = not statistically significant.

between centers that remained for two waves versus three: The percentage of head teachers with a B.A. degree was significantly higher in centers that remained for all three waves.

These findings of substantial differences in initial quality between centers that left the study after one wave and those that remained for at least one additional wave raise questions about the comparability of the sample across waves and the representativeness of the final sample. This attrition also restricted the range of quality in the sample, which may have reduced the likelihood of finding effects of quality on other outcomes. This makes examination of the sample characteristics in Table 5.4 especially important.

Table 5.4 presents child-level data for Wave 1 centers. (Results for Waves 2 and 3 are provided in Tables A.5.1 and A.5.2 in the appendix to this chapter.) Table 5.5 presents the pro-

Table 5.3
Comparison of Centers That Dropped Out After Two Waves with Those That Stayed for All Three

	Dropped Out After Wave 2 (N = 9)		Stayed After Wave 2 (N = 43)		Comparison	
Variable	Mean	SD	Mean	SD	t value	p. value
Mean ECERS-R score	4.67	1.21	5.16	.94	1.37	NS
Number of head teacher ECE credits	9.71	7.55	11.63	9.32	.58	NS
Percentage of head teachers with a B.A. degree	.03	.07	.21	.32	3.26	<.005
Head teacher years of teaching experience	7.62	6.59	8.62	4.91	.53	NS
Years of director administrative experience	8.80	6.96	8.80	5.03	- .00	NS
Percentage of directors with a B.A. degree	.33	.50	.46	.47	.72	NS
Average class ratio	5.55	1.32	5.82	1.43	.53	NS
Percentage accredited	.11	.33	.23	.43	.80	NS

NOTE: The analyses presented above are based on Wave 2 data.

portion of Wave 1 centers with selected characteristics. (Analogous results for Waves 2 and 3 may be found in Tables A.5.3 and A.5.4 in the appendix to this chapter.) Each of the measures included in these tables is described in detail in Chapter Two.

Relationships of the Q-QRIS Components to the Process-Quality Measures

The two process-quality measures, the Caregiver Interaction Scale and the Pre-Kindergarten Snapshot, were included in the study as means of assessing how well the star ratings and the individual Q-QRIS components were measuring quality. We did not expect perfect correlations. The CIS and Pre-K Snapshot look at very specific aspects of classroom process, while the Q-QRIS components focus on several different aspects of structural quality, and the star ratings average across components. Only the ECERS-R directly assesses process, and process constitutes just a portion of a measure, which focuses on materials, health, and safety. However, since structural quality characteristics are thought to drive process quality, we did expect to see relationships among the Q-QRIS components and the process-quality measures.

Table 5.6 provides all of the correlations at Wave 1 between the Q-QRIS components and the process-quality measures at the classroom level. Given the very small sample size, we were not able to conduct regressions that would have allowed us to control for key covariates. Instead, we simply compared the correlation coefficients across waves. Wave 1 results were selected for presentation because they are based on the largest sample available to us in these analyses. (Results for Waves 2 and 3 are available in Tables A.5.5 and A.5.6 in the appendix to this chapter.)

The data reveal a small number of significant correlations between the Q-QRIS components and the CIS. Nearly all were in the expected direction. Higher ECERS-R score was associated with lower punitiveness, less detachment, and a more positive relationship. The FP parent points also relate well to the CIS. The FP parent survey showed the same relationship to the process-quality measures as the ECERS-R: significant negative correlations with

Table 5.4
Descriptive Statistics for Wave 1 Centers

Variable	Mean	SD	Min	Max	Alpha
Child-Level Data (N = 1368)					
Woodcock Johnson					
Letter Word Identification summary score	100.086	17.658	63	181	
Applied Problems summary score	97.224	14.600	54	140	
Passage Comprehension summary score	119.110	13.134	73	171	
PPVT Standard Score	91.222	15.080	20	147	
Classroom Behavior Inventory					
Apathy subscale	2.271	0.768	1	5	.73
Considerateness subscale	3.402	0.882	1	5	.89
Creativity/curiosity subscale	3.558	0.832	1	5	.91
Dependence subscale	2.497	0.803	1	5	.80
Distractibility subscale	2.692	0.882	1	5	.86
Hostility subscale	2.533	1.076	1	5	.88
Independence subscale	3.626	0.719	1	5	.85
Intelligence subscale	3.292	0.910	1	5	.95
Task-orientation subscale	3.268	0.875	1	5	.93
Child's age at assessment (months)	47.338	8.504	27.64	72.70	
Average family income	45,400.18	38,120.49	2,500	250,000	
Average hours per week child attends provider	29.570	11.790	3	85	
Average months child has attended provider	14.190	10.472	0	72	
Child-Rearing Practices Report					
Nurturance subscale	5.366	0.428	1	6	.87
Restrictive subscale	3.167	0.617	1.16	5.89	.82
Classroom-Level Data (N = 148)					
Average class ratio	6.206	1.981	2.26	12	
ECERS-R class mean	4.436	1.296	1.79	6.77	
Head teacher ECE credits	12.941	15.401	0	71	
Head teacher teaching experience (years)	7.639	6.135	0	36	
Provider-Level Data					
Director years of administrative experience (N = 65)	6.55	5.50	0	24	
Director ECE credits	23.599	13.048	0	6.77	
FP parent survey (version 4) (N = 50)	21.75	6.295	5	28	
FP provider survey (version 4) (N = 50)	26.85	4.320	15	31	

the CIS detachment and punitiveness subscales and a positive correlation with the CIS positive relationship subscale. FP provider responses were also correlated negatively with the CIS detachment subscale and positively with the positive relationship subscale. However, neither the ECERS-R nor the FP measures correlated with the Pre-K measure of process quality.

Two measures of staff credentials—director B.A. and head teacher ECE credits—are significantly associated with a number of process-quality measure subscales. Director B.A. correlated negatively with the CIS punitiveness subscale and positively with the permissiveness subscale. Head teacher ECE credits were negatively correlated with the detached subscale.

Table 5.5
Wave 1 Data on Selected Child, Classroom, and Center Characteristics (Percentages)

Variable	Mean	SD	Min	Max
Child-Level Data (N = 1368)				
Parent has a B.A. degree	16	30	0	100
Family speaks English only	89	25	0	100
Child has learning problems	18	32	0	100
Child is of a minority group	42	41	0	100
Child is female	50	50	0	100
Child attends provider full-time	83	38	0	100
Classroom-Level Data (N = 148)				
Head teacher has an A.A. degree	43	49	0	100
Head teacher has a B.A. degree	12	31	0	100
Provider-Level Data (N = 65)				
Director has B.A. degree	42	47	0	100
Head Start	20	40	0	100
Nonprofit	65	48	0	100
Low intervention level	17	38	0	100
Medium intervention level	71	46	0	100
High intervention level	12	33	0	100
Accreditation status	11	31	0	100

Table 5.6
Wave 1 Correlations Between Q-QRIS Components and Process Variables

	Q-QRIS Components									
Process Variable	Average Class Ratio	ECERS-R Class Mean	FP Parent Survey	FP Provider Survey	Director B.A. Degree	Director Admin Exp	Head Teacher B.A. Degree	Head Teacher ECE Credits	Head Teacher Teaching Exp	Accred-itation Status
Caregiver Interaction Scale										
Detachment	0.09	−0.27*	−0.54*	−0.40*	−0.01	0.03	0.11	−0.26*	0.01	0.03
Punitiveness	0.19	−0.47*	−0.32*	−0.19	−0.28*	−0.11	0.04	−0.12	−0.20	−0.16
Permissiveness	−0.00	0.22	−0.04	0.00	0.34*	0.19	0.03	−0.22	−0.03	0.28*
Positive relationship	−0.22	0.40*	0.55*	0.39*	0.17	0.07	0.03	0.15	−0.10	0.05
Pre-K Summary Scores										
Peer play	−0.03	−0.15	−0.09	−0.19	−0.07	0.12	−0.08	−0.24	−0.13	0.11
Nonplay	−0.02	−0.04	0.01	−0.01	0.04	0.08	−0.07	0.08	0.11	−0.10
Cognitive play	−0.12	−0.01	−0.05	−0.11	0.09	0.01	−0.07	−0.38*	−0.09	0.25
Adult involvement	−0.23	0.24	0.30	0.12	0.35*	0.14	0.12	0.02	−0.08	0.16

* Significance = p. < .025.

Head teacher ECE credits correlated negatively with cognitive play subscales, one of the few unexpected correlations. Director B.A. correlated positively with the Pre-K adult involvement subscale. Accreditation correlated positively with the permissive subscale of the CIS. Ratios were not correlated with either of the process-quality measures.

The pattern of findings found in the Wave 1 data was generally not observed in Waves 2 and 3. Across waves, the typical pattern was a significant correlation at Wave 1, for which the sample size was largest, and correlations approaching zero at Wave 2 and Wave 3. On a few occasions, we found not just a lessening of the significance of the results, but inconsistent results.

This pattern of findings, in which the more process-focused Q-QRIS components relate to some of the CIS subscales but hardly at all to the Pre-K at Wave 1, and the correlations decline and sometimes reverse in later waves, presents a decidedly mixed picture about the functioning of the Q-QRIS. Next, we analyze the relationship between star ratings and the process-quality measures to see whether these patterns hold.

Relationship of the Star Ratings to the Measures of Process Quality

Because the process-quality measures were collected in just one preschool classroom, there is no within-center variability on these measures, and sample sizes for these analyses are small. Consequently, we could not control for child and program background characteristics in our analyses. We ran separate analyses for each wave, and we report Wave 1 data here because of larger sample sizes.

As shown in Table 5.7, star rating was significantly related to two CIS subscales, detachment and positive relationships. Both relationships were in the predicted direction and, in both cases, 4-star providers scored best, although the relationships were not completely linear for either subscale. There were no significant relationships between star rating and the Pre-K.

Table 5.7
Relationship of Process-Quality Measures to Star Ratings in Centers: Wave 1

	Mean						
	Star 0	Star 1	Star 2	Star 3	Star 4	F	df
Caregiver Interaction Scale							
Detachment	6.951	7.350	5.711	5.163	4.818	3.36*	4,60
Permissiveness	9.208	8.517	9.158	9.353	9.851	0.64	4,60
Positive relationship	27.518	29.077	32.818	34.396	35.367	6.78*	4,60
Punitiveness	12.488	12.011	10.090	9.992	8.879	2.57	4,60
Pre-Kindergarten Snapshot							
Responsive adult involvement	0.333	0.360	0.433	0.402	0.789	2.66	4,60
Peer play	2.379	2.241	2.321	2.247	2.232	0.98	4,60
Nonplay	1.951	1.190	1.443	1.503	2.958	1.09	4,60
Cognitive play	3.148	3.110	3.077	3.225	2.889	.54	4,60

NOTE: df = degrees of freedom.

* Significance = p. < .025.

There were no significant relationships in Wave 2 or Wave 3 data, which may reflect, at least in part, smaller sample sizes. While 4-star providers scored best on several subscales in Wave 2, the results were not in the expected direction, and the patterns of findings were not replicated in Wave 3. (See Tables A.5.7 and A.5.8 in the appendix to this chapter for Wave 2 and Wave 3 results.)

Examining both the individual components and the star ratings, we see few indications in the Wave 1 data of associations between the Q-QRIS and the process-quality measures. Furthermore, these findings do not hold up in the smaller subsequent waves of data.

Relationships Between Q-QRIS Components and Child Outcomes

Examining the relationship between the Q-QRIS components and child outcomes was another way to assess the validity of the Q-QRIS. As presented in the logic model in Chapter One, the theory behind QRISs posits that quality care promotes improved child outcomes.

As described in Chapter Two, a variety of child outcome measures were utilized in the study. They are briefly listed below as a reminder to readers (see Table 5.8). Efforts to create separate cognitive and noncognitive scores through factor analysis were not successful, so the 13 outcomes (plus five more in Wave 3 only) are treated individually.

The child outcomes and the multiple Q-QRIS component measures resulted in a large number of significance tests; the number of test increased when we added the SDQ in Wave 3.

Table 5.8
Child Outcome Measures

Woodcock-Johnson
Letter-Word Recognition
Passage Comprehension
Applied Programs
Peabody Picture Vocabulary Test
Child Behavior Inventory
Apathy
Considerateness
Creativity
Dependence
Distractibility
Hostility
Independence
Intelligence
Task orientation
Strengths and Difficulties Questionnaire (Wave 3 only)
Emotional symptoms
Conduct problems
Hyperactivity and inattention
Peer relationship problems
Pro-social behavior

For this reason, we adopt a significance level of p. < .025 to minimize the proportion of false positives.

Our data reveal virtually no relationships between the Q-QRIS components and child outcomes. Below we focus on Wave 1 cross-sectional data because that wave had the largest sample. Only a handful of relationships were significant, as shown in Table 5.9. However, these significant results are not replicated, for the most part, in the cross-sectional data from Waves

Table 5.9
Cross-Sectional Regressions: Wave 1 Q-QRIS Components Predicting Wave 1 Child Outcomes

		Q-QRIS Components									
Child Outcomes		Average Class Ratio	ECERS-R Class Mean	FP Parent Survey	FP Provider Survey	Director B.A. Degree	Director Admin. Exp.	Head Teacher B.A. Degree	Head Teacher ECE Credits	Head Teacher Teaching Exp	Accred-itation Status
WJ – LWI	*Est.*	.042	.040	–.066	.015	.101	.005	–.168	.000	.008	–.092
	SE	.041	.057	.070	.063	.158	.008	.122	.003	.007	.139
WJ – PC	*Est.*	.011	.059	.006	–.012	.120	–.009	–.028	.001	–.000	–.056
	SE	.029	.036	.054	.039	.104	.005	.084	.002	.004	.095
WJ – AP	*Est.*	.041	.072	–.020	–.026	–.061	.012	–.092	–.001	–.007	–.160
	SE	.034	.047	.062	.049	.125	.007	.102	.003	.005	.120
PPVT	*Est.*	.053	.068	.032	–.015	.112	.006	–.058	–.002	–.006	–.003
	SE	.036	.051	.059	.057	.153	.008	.096	.003	.006	.127
CBI – Apathy	*Est.*	–.026	–.013	–.082	–.019	.147	–.006	–.063	–.001	.012	–.085
	SE	.044	.069	.109	.098	.201	.010	.124	.003	.006	.175
CBI – Considerate	*Est.*	.007	.037	–.048	–.018	.050	–.015	.028	.001	–.000	.311
	SE	.042	.066	.085	.063	.162	.008	.113	.003	.006	.143
CBI – Creative	*Est.*	.009	.049	.026	.027	–.030	–.006	.177	.001	–.015*	.109
	SE	.054	.068	.107	.090	.245	.010	.116	.003	.006	.176
CBI – Depend	*Est.*	–.043	.043	–.030	.016	–.068	.010	.141	–.000	.003	–.111
	SE	.051	.074	.133	.097	.240	.012	.130	.003	.006	.198
CBI – Distract	*Est.*	–.035	–.043	–.034	.039	–.006	.002	–.111	–.002	.001	–.018
	SE	.044	.065	.111	.078	.201	.010	.127	.003	.006	.180
CBI – Hostile	*Est.*	–.013	–.078	.057	.020	.064	.019*	–.100	–.001	–.003	–.385
	SE	.039	.063	.072	.068	.152	.008	.114	.003	.006	.145
CBI – Independent	*Est.*	–.023	.015	.007	.020	.087	–.012	.002	.003	–.003	.161
	SE	.055	.073	.115	.096	.216	.010	.119	.003	.006	.181
CBI – Intelligent	*Est.*	–.050	.050	–.033	.029	–.162	–.006	.045	–.002	–.002	.049
	SE	.050	.069	.090	.075	.230	.010	.107	.003	.006	.170
CBI – Task-Oriented	*Est.*	–.007	.045	–.019	.011	–.020	–.014	–.047	.002	.003	.067
	SE	.046	.066	.102	.075	.217	.010	.119	.003	.006	.164

NOTES: *Est.* = parameter estimate; SE = standard error.

* Significance = p. < .025.

2 and 3. (See Documents A.5.1 and A.5.2 in the appendix to this chapter for these Wave 2 and 3 results.)

Sensitivity Analyses

This cross-wave inconsistency might be explained by children's limited exposure to a given provider. One would not expect quality to have an effect unless children have had substantial opportunity to be influenced by that quality; for this reason, the length of time children had attended their current provider was always included as a covariate in our models. Sensitivity analyses were conducted to determine whether links between child-care quality and child functioning varied for children who experienced a high "dosage" of care at their center based on the length of time they had attended that provider and the number of hours per week they spent in care.

We also tested whether the link between child-care quality and child functioning varied for children who came from lower-income families. Specifically, separate models were run for children whose family income fell below the median for this sample ($37,500), and children whose family income fell below the 25th percentile for this sample ($17,500). The sensitivity analyses necessarily relied on subsets of the study sample, so the number of children included in each group was small.[4] Given the small sample sizes, it is not surprising that these analyses produced few effects. These results do suggest that the relationships between child-care quality and child outcomes are not stronger for low-income children or for children with stronger doses of care.

Lagged Analysis

Lagged analyses allow us to explore whether there is a delay in the effects of the Q-QRIS. Two sets of lagged analysis results were available: one set that predicts child outcomes in Wave 2 from Q-QRIS components in Wave 1 and another that predicts child outcomes at Wave 3 based on Wave 2 Q-QRIS scores. In the discussion below, we focus on Wave 1 to Wave 2 because the sample size was larger.

Although there were some significant associations, the total number—four—is very small, as shown in Table 5.10. Most Q-QRIS components bore no significant relationships to child outcomes. We found, for example, that Wave 1 child-staff ratios did not predict any of the Wave 2 child outcomes examined in the current study. There were also no relationships between ECERS-R or accreditation and any child outcome, and just one among all the training and education variables. However, the FP parent survey (version 4 at Wave 1) predicted higher levels of considerateness among children on the CBI one year later ($p. < .025$). The FP parent survey also negatively predicted CBI hostility scores ($p. < .025$).

In contrast to the largely positive findings for the FP parent survey above, the association between the FP provider survey and study outcomes was counterintuitive. The FP provider survey was a positive predictor of hostility on the CBI two years later ($p. < .025$).

Comparison across the two sets of lagged analyses (i.e., Wave 1 to Wave 2 and Wave 2 to Wave 3) revealed inconsistent patterns. While a number of predictors were significant within

[4] For example, in Wave 1 (where Ns are largest), the number of children in the below-the-median income group ranged from 583 to 610; the number in the group with a family income below the 25th percentile for this sample was 401 to 425; the number of high-dosage children ranges from 272 to 298. Ranges are given because sample size varied slightly across analyses as a result of multiple imputation.

Table 5.10
Lagged Regressions: Predicting Wave 2 Child Outcomes from Wave 1 Q-QRIS Components

		Q-QRIS Components									
Child Outcomes		Average Class Ratio	ECERS-R Class Mean	FP Parent Survey	FP Provider Survey	Director B.A. Degree	Director Admin. Exp.	Head Teacher B.A. Degree	Head Teacher ECE Credits	Head Teacher Teaching Exp	Accred-itation Status
WJ – LWI	*Est.*	.033	–.041	.005	.001	–.045	.006	.145	.005	–.002	–.033
	SE	.032	.072	.020	.018	.149	.011	.259	.004	.008	.184
WJ – PC	*Est.*	.032	.064	.045	–.032	–.158	–.007	–.344	–.000	.007	–.182
	SE	.033	.069	.020	.018	.152	.011	.259	.004	.009	.171
WJ – AP	*Est.*	–.008	.033	.018	–.007	–.004	–.002	.136	–.006	–.002	–.071
	SE	.031	.076	.019	.017	.129	.011	.253	.004	.008	.173
PPVT	*Est.*	–.020	.060	.037	–.029	–.053	–.005	.132	.003	–.002	.012
	SE	.031	.067	.018	.016	.131	.010	.237	.004	.008	.170
CBI – Apathy	Est.	–.011	–.029	–.054	.049	.196	.014	–.037	.001	.005	–.040
	SE	.048	.092	.031	.027	.232	.018	.317	.005	.011	.287
CBI – Considerate	*Est.*	.024	.044	.053*	–.044	–.069	–.013	.128	.002	–.019	–.124
	SE	.038	.076	.023	.021	.174	.015	.280	.005	.010	.216
CBI – Creative	*Est.*	–.005	–.094	.035	–.033	.040	–.003	.092	–.000	.001	–.016
	SE	.045	.101	.028	.025	.208	.017	.326	.005	.012	.279
CBI – Depend	*Est.*	–.010	–.028	–.034	.033	.072	.026	–.697	–.008	–.006	–.174
	SE	.050	.101	.033	.029	.261	.020	.318	.006	.013	.316
CBI – Distracted	*Est.*	–.009	–.074	–.040	.038	.043	.029	–.451	–.014*	.019	–.336
	SE	.042	.085	.026	.023	.207	.016	.319	.005	.011	.251
CBI – Hostile	*Est.*	.014	.065	–.074*	.060*	.352	.007	–.123	–.010	.012	–.045
	SE	.038	.084	.024	.021	.178	.014	.271	.004	.010	.224
CBI – Independent	*Est.*	–.025	–.024	.031	–.035	.233	–.020	.427	.006	–.010	.312
	SE	.044	.096	.028	.025	.197	.016	.331	.006	.011	.271
CBI – Intelligent	*Est.*	.011	–.078	.030	–.030	.075	.002	.274	.001	–.002	.151
	SE	.041	.087	.025	.022	.195	.016	.303	.006	.011	.248
CBI – Task-Oriented	*Est.*	–.003	.014	.013	–.021	.144	–.011	.496	.003	–.014	.100
	SE	.039	.088	.024	.021	.178	.015	.302	.005	.011	.231

NOTES: *Est.* = parameter estimate; SE = standard error.

* Significance = p. < .025.

analyses, none of the same predictors were significant across both sets of analyses. (See Document A.5.3 in the appendix to this chapter for Wave 2 to Wave 3 lagged analyses.)

These findings reveal few relationships between the Q-QRIS components and child outcomes; the lack of relationships is apparent whether the data are examined cross-sectionally or over time. Interestingly, the FP measure, the Q-QRIS component that received the most conceptual and empirical attention in this study, was one of the best predictors of child outcomes. However, these relationships were only significant for Waves 2 and 3, the same waves for which

the FP did not relate to the other Q-QRIS components. Additionally, the two parts of the measure—the parent and provider surveys—were found in several cases to relate significantly but in opposite ways to child outcomes. Reasons for these findings are not immediately apparent. One explanation may lie in the fact that directors report on their own FP performance in this high-stakes context. Clearly, more empirical work needs to be done with the FP.

Relationships Between Star Ratings and Child Outcomes

Star ratings are, of course, the ultimate output of the Qualistar QRIS. We regressed child outcomes on star rating levels, controlling for a range of child and program characteristics (see a listing of these variables in the methods section of this chapter).

We examined the associations between star ratings and child outcomes both cross-sectionally and over time. In each case, we compared child outcomes at a given star rating, e.g., 3 stars, against outcomes on the same measure at 0 stars. As shown in Table 5.11, we found just one significant association between star ratings and child outcomes in our Wave 1 data, which are presented here because of the larger sample sizes in Wave 1. (See Tables A.5.9 and A.5.10 in the appendix to this chapter for Wave 2 and Wave 3 results.) The single significant finding in Wave 1 was an association between a 2-star rating and lower levels of apathy as measured on the Child Behavior Inventory (CBI). Waves 2 and 3 also showed very few significant relationships. Furthermore, some findings were counterintuitive, and patterns of significance did not replicate across waves.

Lagged analyses, in which star ratings in the previous wave predict child outcomes in the next wave, were also run because it seemed possible that it would take time for a given star rating to affect child outcomes. These analyses would allow us to determine if star ratings in an earlier wave predicted child outcomes in a later one. The lagged analyses showed very few effects of prior star ratings: The Wave 1–Wave 2 analyses revealed one significant association: a significant positive association between a 3-star rating in Wave 1 and PPVT score in Wave 2. There was a single statistically significant association between Wave 2 star ratings and Wave 3 child outcomes: This association was between a 2-star rating in Wave 2 and PPVT score in Wave 3. (See Tables A.5.11 and A.5.12 in the appendix to this chapter for both tables.)

If child-care quality indexed by star ratings is having an effect on children's outcomes, one would look not only for significant associations, as discussed above, but also for patterns of associations in which child outcomes improved linearly as star rating increased. If children in 3-star settings do better overall than children in 2-star settings, and children in 4-star settings do best, this pattern suggests that quality, as indexed by star rating, is associated with child outcomes even if the effects may not be strong and statistically significant. This linear pattern occurred only twice in these analyses, suggesting the absence of a meaningful association between star ratings and child outcomes.

Examining the relationships between child outcomes and both the individual Q-QRIS components and the star ratings, we found very few relationships. We discuss possible reasons for the lack of associations at the end of this chapter.

Table 5.11
Cross-Sectional Regressions: Wave 1 Star Ratings Predicting Wave 1 Child Outcomes

Child Outcomes		Star 1	Star 2	Star 3	Star 4
WJ – LWI	*Est.*	–0.068	–0.467	–0.250	–0.431
	SE	0.215	0.230	0.237	0.510
WJ – PC	*Est.*	0.002	0.037	0.049	–0.273
	SE	0.152	0.144	0.152	0.349
WJ – AP	*Est.*	–0.088	–0.206	–0.115	0.089
	SE	0.182	0.195	0.205	0.462
PPVT	*Est.*	0.198	0.061	0.275	0.392
	SE	0.191	0.201	0.200	0.440
CBI – Apathy	Est.	–0.121	–0.597*	–0.450	–0.633
	SE	0.228	0.248	0.254	0.595
CBI – Considerate	*Est.*	0.105	–0.142	–0.253	–0.798
	SE	0.197	0.214	0.219	0.509
CBI – Creative	*Est.*	0.459	0.491	0.407	0.521
	SE	0.241	0.261	0.262	0.628
CBI – Depend	*Est.*	0.014	–0.147	–0.133	0.651
	SE	0.256	0.279	0.279	0.668
CBI – Distracted	*Est.*	–0.351	–0.473	–0.141	–0.398
	SE	0.214	0.232	0.236	0.555
CBI – Hostile	*Est.*	–0.116	–0.051	0.088	0.877
	SE	0.199	0.216	0.223	0.517
CBI – Independent	*Est.*	0.398	0.145	0.100	0.628
	SE	0.230	0.250	0.254	0.601
CBI – Intelligent	*Est.*	0.182	0.082	–0.024	0.706
	SE	0.214	0.230	0.232	0.553
CBI – Task-Oriented	*Est.*	0.292	0.183	0.048	0.290
	SE	0.205	0.223	0.227	0.532

NOTES: *Est.* = parameter estimate; SE = standard error.

* Significance = p. < .025.

Relationships Between Process-Quality Measures and Child Outcomes

The two process-quality measures, the Caregiver Interaction Scale and the Pre-K Snapshot, were included in the study as part of the effort to validate the Q-QRIS. One aspect of that validation involved examining the relationships between the process-quality variables and child outcomes, and comparing the strength and direction of those relationships with those found between the Q-QRIS components and child outcomes.

To assess these relationships, we ran regressions similar to those we used in assessing the relationship between the Q-QRIS components and child outcomes (discussed above). These regressions controlled for provider and family characteristics.

The four subscales of each of the process measures bore very few relationships to the child outcomes. On the CIS, none of the tests were significant; on the Pre-K, two achieved statistical significance (p. < .025). (See Documents A.5.4–A.5.9 in the appendix to this chapter for these data.) The significant relationships between the Pre-K and child outcomes focused on the adult involvement subscales of the Pre-K and the CBI. Pre-K adult involvement was negatively associated with CBI apathy and positively associated with CBI creativity in Wave 1. Results for Waves 2 and 3 showed some significant associations, but these differed by wave and were counterintuitive at times.

The very limited number of statistically significant relationships between the process-quality measures and the child outcomes mirrors the pattern found for the relationship between the Q-QRIS components and child outcomes. There are a number of reasons for the lack of associations, discussed below.

Conclusion

The findings reported in this chapter suggest that the star ratings and the individual Q-QRIS show few associations that are consistent across waves to the process-quality measures. Neither the star ratings, the individual Q-QRIS components, nor the process-quality measures relate strongly to the child outcome measures.

A number of factors may explain why stronger relationships were not found between the star ratings and the individual Q-QRIS components and the criterion and child outcome measures. The most important of these are presented below.

Provider Self-Selection. Providers in participating HB 1297 communities were eligible to join this study; we had no control over the application process. As a result, we do not know whether and to what extent self-selection bias may have resulted in a sample of providers that differ from randomly selected providers in terms of their characteristics or those of the children they serve. Certainly, it is reasonable to assume that these providers are more motivated to improve their quality. Perhaps their somewhat higher quality (as measured by the ECERS-R) in Wave 1 contributed to the lack of observed relationships.

Child Sample. We did, of course, control for a large number of family background variables that are most likely to affect results in analyses. However, controls for background factors can never equal the power of strong design elements, such as random assignment. Consequently, a number of unobserved characteristics may be muting potential effects.

Center Attrition. While the rate of center attrition was not high overall, center attrition was not random. The centers that dropped out of the study contained classrooms that had lower ECERS-R scores at Wave 1, as discussed above. These differences were in some cases substantial. For example, centers that left the study after Wave 1 had ECERS-R scores that were more than a full point lower on a 7-point scale than those that remained in the study. Similarly, the family child-care providers that did not remain in the study through Wave 3 were of lower quality at Wave 1 on the FDCRS, although the quality differences between homes that stayed in and those that left the study were much smaller than in the case of the centers and were only rarely statistically significant (see Chapter Six).

These nonrandom dropout patterns suggest that the analyses for each wave of data are based on slightly different samples. This fact likely contributes to the lack of consistency in the findings when cross-sectional analyses are compared.

Staff-Retention Rates. Low levels of compensation in the child-care sector contribute to high turnover rates among child-care staff. While pay has improved to some degree in some locations, for many staff, low pay, the lack of any career ladder, long hours, and demanding work are the reality of working in the child-care sector. Not surprisingly, turnover rates in our sample were high. Staff turnover reduces comparability across waves. Six months following the Wave 1 assessments, directors in 43 centers were asked how many staff had left the center in the past *three* months for any reason. The average response was 2.16 (it was 1.76 and 1.14 for Waves 2 and 3, respectively). Given that in Waves 2 and 3 centers had an average of 10.39 and 10.69 staff, respectively (this includes floaters and teaching assistants), it appears that these centers experienced substantial staff turnover.

Child Attrition. Fewer than 10 percent of the children in Wave 1 were available for assessment at Wave 3. Because of this enormous attrition, it was necessary to refresh the sample in Waves 2 and 3. This may have introduced bias across waves, since each wave of data consisted of a large number of nonoverlapping children. Of course, this issue would not affect cross-sectional analyses. As with center attrition, having nonoverlapping samples across waves may explain some of the inconsistencies in findings across waves.

Movement. Movement of children and staff within a center for nonprogrammatic reasons has been virtually ignored in the child-care literature. Programmatic movement refers to movement that serves an educational function, such as a staff member taking children to a music class; programmatic moves are generally considered to be good practice. Stability of staff assignments to classrooms and stability of staff in classrooms over the course of the day are implicitly assumed in analyses that relate staff characteristics, particularly staff training and education, to child outcomes. Our data suggest that there is a significant amount of staff movement between classrooms over the course of a day, much of it nonprogrammatic. For example, in response to an item on the Organizational Change Survey, 26 percent of directors indicated that they move staff in order to adhere to ratio requirements (i.e., not for programmatic reasons) one or more times per day. Another 19 percent indicated they do so weekly. Given the strong social desirability factor at play in this context and the fact that the Q-QRIS has high stakes associated with it, these figures likely underestimate the true rate of nonprogrammatic movement. Failure to capture that movement in analyses of the effect of staff characteristics on child outcomes muddies those assessments: Measures of staff characteristics should reflect the actual characteristics of the staff who interact with the children in a classroom most of the time.

Are We Employing the Right Measures?

Our measures of child-care quality are very broad. Many of them, such as ratios, are structural proxies that are quite distal from the child outcomes we hope to affect. For example, higher letter-word recognition scores on the Woodcock-Johnson are facilitated by a rich verbal environment, but our measures of that environment are limited. Most of the measures of quality that we and others use—even those that examine process characteristics—do not actually assess the verbal environment (with the exception of limited items on the ECERS-R). We assume that programs with better ratios and higher ECERS-R scores provide a richer verbal environment than lower-scoring programs, but there is limited empirical support for these assumptions. If we determined that letter-word recognition is an important skill for children

to acquire and therefore should be a goal of child-care programs, we might want to measure aspects of the verbal environment more directly. This may increase the likelihood of finding effects and would provide incentives for programs to pay attention to and improve that environment.

The early childhood education research agenda has been influenced by an emphasis on preparing children to succeed at school. Such a focus (especially if supported by empirical data) may provide a useful tool for advocates working toward funneling greater resources into educational settings for our youngest children. However, it may introduce a focus that is not ideal for young children, where measures of such things as growth, satisfaction, and self-esteem may be more important than pre-academic skills (National Research Council and Institute of Medicine, 2000a). However, it is important to note that we did not find associations in this study between program quality and the less-academic outcomes we employed, such as distractibility, task orientation, and conduct problems. The strongest effects of early intervention programs are reported in a handful of studies that focused on very low-socioeconomic-status samples, used random assignment, relied on a strong, consistent intervention, and compared the high-quality care to no care. Outside of studies using such designs with these samples, the limited effects we found may, in fact, be more typical than not. Perhaps expecting to find effects on measures of child cognition, peer relationship problems, or pro-social behavior is expecting too much.

It may make more sense, particularly until we can build a stronger empirical basis for our measures of quality, to focus on program outputs rather than child outcomes. We might focus on what children are doing in child care, particularly their engagement in developmentally appropriate tasks in a safe and supportive environment.

Sample Size Issues. As noted previously, attrition of providers and children was substantial. Even with the addition of new children at Waves 2 and 3, sample size limited the sorts of analyses we could do. In many analyses, we had to rely on correlations, which meant that we could not control for key variables. In other analyses, we relied on imputation as well. Sample size in the data on home providers was a known problem from the beginning, but Qualistar Early Learning felt that it was important to have at least some data about the functioning of the Q-QRIS in family child-care settings. Small sample size forced us to rely exclusively on correlational analyses for these data, which are presented in Chapter Six.

Some of the above methodological issues are specific to the current study (e.g., child attrition rates). Some reflect the reality of conducting research on child care (e.g., staff turnover). Together they contribute to the inconsistent and limited relationships we found between child-care quality as captured by the Q-QRIS components and the process-quality measures and measures of child functioning.

CHAPTER SIX

Family Child-Care Providers

Background

While most children cared for outside their homes are enrolled in child-care centers, many children receive out-of-home care in family child-care homes. A recent study reported that 14 percent of working mothers use family home care as their primary form of child care for their children age six or under (Boushey and Wright, 2004). Johnson (2005) reports that almost one-quarter of children spend some time in family child care. These homes are less likely to be included in research studies and only rarely are the central research focus. One important reason is that the cost per child of conducting such research is high, since most homes serve very few children.

But the sheer numbers of children in such care demands some study of the care provided in these homes and the effects of that care on the children served (Kontos, Howes, and Galinsky, 1996). After a brief review of research on quality in family child-care homes, we describe the family child-care provider sample examined in the current study and present analyses of these data.

Research on Quality of Care in Family Child-Care Homes

Several studies have found that the quality of care children receive in family child-care homes is associated with their social, cognitive, and emotional functioning. After controlling for the requisite family background characteristics, higher levels of process and structural quality have been associated with higher levels of cognitive (Goelman and Pence, 1994; Clarke-Stewart et al., 2002; Kontos, 1994) and social development (Kontos, 1994; Kontos, Hsu, and Dunn, 1994; Clarke-Stewart et al., 2002; NICHD ECCRN, 2003) and with secure attachment to the caregiver (Kontos, Howes, and Galinsky, 1996; Elicker, Fortner-Wood, and Noppe, 1999).

Research Linking Structural Quality Characteristics to Process Quality

Staff training and education have been examined in relation to the family child-care provider's process quality. Several studies have found that the provider's level of formal education is linked to process quality (Burchinal, Howes and Kontos, 2002; Clarke-Stewart et al., 2002; Raikes, Raikes, and Wilcox, 2005). Training has also been found to be a significant predictor of quality (e.g., Whitebook et al., 2004). However, two studies do not replicate this link for formal education (Pence and Goelman, 1991; Doherty et al., 2006) and one found no relation between training and quality (Kontos, 1994). Findings about the relationship between years

of provider experience and quality are inconsistent, with some studies supporting such a link (Stuart and Pepper, 1988), some finding a negative relationship (NICHD ECCRN, 1996; Burchinal et al., 2002), and still others finding no relationship at all (Clarke-Stewart et al., 2002; Kontos, 1994).

Findings about the effect of child-staff ratios on quality in family child-care providers have been similarly inconsistent. Some studies report higher quality levels when child-staff ratios are lower (Clarke-Stewart, Gruber, and Fitzgerald, 1994; Howes and Norris, 1997), whereas some report lower quality when child-staff ratios are lower (Pence and Goelman, 1991; Kontos, Howes, and Galinsky, 1996). Burchinal et al. (2002) reported no relationship at all. Helburn, Morris, and Modigliani (2002) found that in Kontos, Howes, and Galinsky's (1996) family child-care and relative care sample the highest-quality providers actually cared for a larger number of children (and had higher child-staff ratios); these providers typically considered themselves professionals running their own businesses.

Other potential predictors of process quality that have been examined are provider work conditions, intentionality, use of supports, and children's ages. Doherty et al. (2006) provide a good summary of results of these various provider characteristics.

The Family Child-Care Home Sample

Because of the large numbers of children cared for in homes, Qualistar insisted that its Q-QRIS be assessed in homes as well as centers. However, developing a quality rating system that is valid across settings is challenging considering that inherent differences between family child-care providers and centers must be factored into quality measures and attempts to create a summary score that is valid across settings. For example, a QRIS may include points for the director's educational experience in addition to teachers' on a training and education component. What happens in family child-care settings where there normally is only one caregiver? Should a provider's educational background count twice—as a provider and as an administrator? Similarly, empirical analyses comparing the measures of the classroom environment (e.g., the ECERS-R and FDCRS) are not available, making it difficult to fold the two measures into a parallel total quality score.

The child-care homes included in this study also posed a set of problems inherent in conducting research of any sort in this setting: a higher likelihood that these small businesses would close before the study's end.[1] Another problem is the need to collect data in locations where, at best, there would be five or six eligible children and more likely two or three. Perhaps most troubling, it was not possible, given the resources available, to collect data from enough homes to enable those data to be analyzed using sophisticated, multivariate techniques. Data from family child-care providers would provide an interesting contrast to the data from centers. However, power calculations based on fairly conservative assumptions about likely changes in provider Q-QRIS ratings and in children's outcomes revealed that the standard deviations for the minimal detectable effect size exceeded our assumption about likely changes in children's outcome scores for a final sample of 30–40 homes. In short, we lacked the power to detect effects in the home sample. Nonetheless, having some preliminary exploratory data

[1] While turnover rates for family child-care providers tend to be comparable to those for child-care center staff (Whitebook et al., 2004), when a family child-care provider quits, the home closes its doors, whereas centers continue to operate.

about family child-care providers was imperative for Qualistar. We present these data in this chapter. (More data may be found in the appendix to this chapter; the location of specific data is indicated in the discussion below.)

Methods

Characteristics of the Home Sample

Thirty-eight family home providers were recruited in Wave 1. Home providers were considered eligible if they were licensed, served children 2.5 to 5 years of age, and had not previously been assessed on the FDCRS (Harms and Clifford, 1989). Providers had to care for at least one child within the identified age range in Wave 1 to qualify for study inclusion. (See Tables A.6.1 and A.6.2 in the appendix to this chapter for descriptive statistics for Wave 1 homes.) Once included in Wave 1, providers remained in the study even if they no longer served any children within the identified age range in Wave 2 and/or 3.

These 38 providers cared for a total of 123 children that were included in the evaluation in Wave 1. An additional 31 non-evaluation children were also included in Wave 1.[2] Twenty-nine homes remained through the study. (See the attrition section below for more detail.)

Analytic Approach

The analyses were designed to parallel to the extent possible the analyses of the center data presented in Chapters Three, Four, and Five. However, with small *n*'s, we had to rely on correlations rather than multivariate analyses. As a result, we could not control for potentially important background characteristics of programs or children.

The analyses addressed the following questions:

1. How closely related are the Q-QRIS indicators to one another?
2. Did the quality of participating providers that remained in the study from Wave 1 to Wave 3 improve?
3. What was the nature of attrition in the home sample?
4. How closely do the Q-QRIS components relate to the measures of process quality?
5. How well do star ratings predict child outcomes?
6. What are the relationships between the individual Q-QRIS component measures and measures of child outcomes?
7. What are the relationships between the process-quality measures and star ratings?

Findings

Correlations Among the Q-QRIS Components

We examined the relationships among the Q-QRIS measures in the family child-care sample by correlating ratios, FDCRS scores, and provider credentials. The only Wave 1 correlation found to be significant was the negative relationship between head teacher (in the case of

[2] Children who did not have parental consent to participate in the study were included in the Pre-K Snapshot evaluations. This was possible because no identifying information was collected.

homes, provider) education and ratio (in the case of homes, the number of children cared for). This correlation suggests that better-educated providers serve fewer children. This association was not significant in the center data, which underscores the greater choices that home care providers have in organizing care.

Improvement Over Time in Family Child-Care Quality

The data on mean scores on Q-QRIS indicators across waves in Table 6.1 are based only on homes for which we had data across all three waves. There were no significant improvements in homes over time, although FDCRS scores did increase by more than a full point from Wave 1 to Wave 3. Small sample size may be reducing the likelihood of finding significant effects. There was also evidence of a trend toward improvement from Wave 1 to Wave 3 on the other indicators. Perhaps over time, if QI efforts continue, these changes may become statistically significant.

Attrition Analysis

Thirty-eight homes joined the study at its inception. Six homes dropped out of the study before or during Wave 2 data collection, and an additional three homes withdrew from the study between Waves 2 and 3; thus, by the end of the study, 29 homes (76 percent) remained. However, seven of the homes present during Wave 2 were not assessed during Wave 2 because of a time lag that occurred as the design of the study changed due to various logistical issues described in Chapter Two. This means that only 66 percent of the original home sample was actually assessed in all three waves.

A comparison of homes that that stayed for one, two, or all three waves revealed only one significant difference on key Wave 1 Q-QRIS measures: The percentage of caregivers with a B.A. degree was significantly higher in homes that remained for all three waves compared with those that dropped out after two waves, as shown in Tables 6.2 and 6.3.

Table 6.3 compares Wave 2 quality indicators for homes that remained in the study and were assessed for two waves against those that remained for all three.

Table 6.1
Changes Over Waves in Key Quality Indicators for the 22 Homes That Remained in the Study for All Three Waves

	Wave 1				Wave 2				Wave 3			
Variable	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
FDCRS class mean	4.32	.61	2.89	5.26	5.19	.81	3.19	6.18	5.43^{+}	.77	3.00	6.48
Head teacher ECE credits	7.93	15.98	0	71.00	10.02	11.24	0	36.00	12.59	11.64	0	36.00
Head teacher teaching experience	10.08	6.98	1.00	25.00	10.34	7.36	.50	30.00	11.88*	7.87	2.00	35.00
Average class child-staff ratio	4.59	.97	2.60	6.06	4.41	1.25	2.33	8.00	4.22	1.06	2.27	6.33

NOTE: The values reported above are based on imputed data.

* Significance = p. < .025.

$^{+}$ Marginal significance = p. < .05.

Table 6.2
Comparison of Homes That Dropped Out After One Wave with Those That Stayed Longer

	Dropped Out After Wave 1 (N = 6)		Stayed After Wave 1 (N = 32)		Comparison	
Variable	Mean	SD	Mean	SD	t value	p. value
Mean FDCRS	3.57	.57	4.13	.78	1.68	NS
Provider ECE credits	3.33	8.16	8.70	18.07	.71	NS
Provider % B.A.	0.17	.41	0.22	.42	.28	NS
Provider teaching experience	4.00	6.20	8.06	7.20	1.29	NS
Average ratio	4.89	1.22	4.95	1.26	.10	NS

Table 6.3
Comparison of Homes That Dropped Out After Two Waves with Those That Stayed for All Three

	Dropped Out After Wave 2 (N = 3)		Stayed (and Were Assessed) After Wave 2 (N = 22)		Comparison	
Variable	Mean	SD	Mean	SD	t value	p. value
Mean FDCRS	4.73	1.45	5.19	0.81	.85	NS
Provider ECE credits	19.33	15.28	10.02	11.24	–1.30	NS
Provider % B.A.	0	0	0.32	0.48	3.13	p. < .001
Provider teaching experience	9.67	6.51	10.34	7.36	.15	NS
Average ratio	4.32	1.27	4.41	1.25	.11	NS

NOTE: The analyses presented above are based on Wave 2 data.

The seven homes that were assessed at Waves 1 and 3 but not Wave 2 were significantly different from those that were assessed in all three waves on several measures. These seven homes had significantly lower FDCRS scores ($p. < .01$), were significantly less likely to be led by a person with a B.A. degree ($p. < .02$), were significantly more likely to be led by a less-experienced person ($p. < .01$), and had significantly higher ratios ($p. < .01$).

Relationship of the Q-QRIS Components to the Process-Quality Measures

Although not statistically or marginally significant, correlations tended to be in the expected direction, with higher quality on the Q-QRIS components associated with higher quality on both the CIS and Pre-K. However, the effects were limited within waves and inconsistent across waves. (The correlations for each wave are presented in Tables A.6.3–A.6.5 in the appendix to this chapter.)

Relationship of the Star Ratings to the Measures of Process Quality

According to the logic model in Chapter Two, if star ratings are operating in the manner expected, the process-quality measures should show a Guttman pattern in which family providers with a rating of 4 perform better than those a rating of 3, providers with a rating of 3 should perform better than providers with ratings of 2, and so forth. To examine if this pat-

tern holds, we examined average scores on each of the process-quality measures at each level of star ratings. We ran separate analyses for each wave, and report Wave 1 results here because of larger sample sizes. (See Tables A.6.6 and A.6.7 in the appendix to this chapter for Wave 2 and Wave 3 results.)

As shown in Table 6.4, there were no significant associations between star ratings and measures of process quality in Wave 1. Lack of significance may be a function of the small sample sizes (38 homes) and the very small number of homes with 4-star ratings (two). Furthermore, means do not generally increase or decline in stepwise fashion in relation to star level. This lack of patterns supports the findings above of no meaningful relationships between star ratings and measures of process quality.

Relationships Between Q-QRIS Components and Child Outcomes

At Wave 1, the only Q-QRIS components that were significantly correlated with child outcomes were the training and education variables. In virtually all cases, higher scores on these variables (denoting more education, training, and experience) were associated with better child outcomes. T&E variables showed a few significant associations with child outcomes, but the effects were scattered across T&E variables such that the pattern across child outcome variables and over time was inconsistent. See Tables A.6.8–A.6.10 in the appendix to this chapter for these results.)

At Wave 3, the only wave in which the SDQ was used, we found two significant correlations with Q-QRIS components. Higher parent survey points were negatively correlated with prosocial behavior, and higher FDCRS scores were associated with more peer relationship problems. Given that we used the SDQ just once and found just two, counterintuitive, significant correlations, it is difficult to draw any definitive conclusions from these data. (These results may be found in Table A.6.11 in the appendix to this chapter.)

Table 6.4
Relationship of Process-Quality Measures to Star Ratings in Family Child Care Providers: Wave 1

	Mean						
	Star 0	Star 1	Star 2	Star 3	Star 4	F	df
Caregiver Interaction Scale							
Detachment	5.769	6.600	5.375	6.000	5.000	.41	4,33
Permissiveness	9.462	10.000	10.250	10.400	9.500	.38	4,33
Positive relationship	30.154	32.400	32.875	31.400	37.000	.57	4,33
Punitiveness	10.769	11.000	9.625	9.000	9.000	.76	4,33
Pre-Kindergarten Snapshot							
Responsive adult involvement	0.555	0.502	0.507	0.548	0.733	.98	4,33
Peer play	2.545	2.177	2.178	2.056	2.437	1.79	4,33
Nonplay	3.206	7.725	7.063	4.100	0.167	.99	4,33
Cognitive play	3.114	2.963	2.795	3.060	3.527	1.4	4,33

NOTES: Because of small sample sizes, analyses do not include controls for child or program background variables. df = degrees of freedom. None of the values in the table are statistically significant.

Relationships Between Star Ratings and Child Outcomes

Given the small number of homes, we ran analyses without controls for child and program background characteristics to assess these relationships. We ran separate analyses for each wave, and report Wave 1 data here because of larger sample sizes. (See Tables A.6.12 and A.6.13 in the appendix to this chapter for Wave 2 and Wave 3 results.)

As shown in Table 6.5, star rating was significantly related to two of the CBI subscales, consideration and independence, and marginally related to the PPVT, task orientation, and intelligence. In all of these cases, there was a general pattern of better child outcomes as star rating increased, although the progression was not always perfect.

In Waves 2 and 3, we found only one significant relationship. There were virtually no instances in which means on the child outcome variables showed a clear linear progression from one star level to the next; this was true even for the one statistically significant relationship.

Relationships Between Process-Quality Measures and Child Outcomes

These relationships were very limited and significant correlations varied across waves. At Wave 3, we found a few more associations between the SDQ and the process measures. Pre-K cognitive play was significantly associated with SDQ conduct problems (r = –.42; p. < .02) and with total SDQ score (r = –.41; p. < .02). Pre-K peer play was significantly and negatively related to SDQ conduct problems (r = –.57; p. < .001) and to SDQ total score (r = –.44; p. < .01). However, none of the significant correlations between the process variables and the child outcomes

Table 6.5
Relationship of Star Ratings to Child Outcomes: Wave 1

Child Outcome	Mean						
	Star 0	Star 1	Star 2	Star 3	Star 4	F	df
WJ – Letter-Word Identification	102.714	100.347	106.671	108.620	110.850	1.42	4,33
WJ – Passage Comprehension	123.447	121.226	124.339	126.753	124.800	0.91	4,33
WJ – Applied Problems	102.912	103.102	109.712	106.276	112.750	1.54	4,33
PPVT	94.541	95.113	98.500	101.530	110.050	2.77+	4,33
CBI – Apathy	2.032	1.881	1.970	1.672	1.575	0.89	4,33
CBI – Considerate	3.398	3.400	3.446	3.827	4.133	3.54*	4,33
CBI – Creative	4.121	3.803	3.732	4.218	4.282	2.04	4,33
CBI – Depend	2.439	2.616	2.512	2.481	2.250	0.3	4,33
CBI – Distract	2.619	2.462	2.618	2.384	1.800	1.55	4,33
CBI – Hostile	2.531	2.689	2.669	2.346	1.850	1.41	4,33
CBI – Independent	3.813	3.553	3.695	4.162	4.317	3.21*	4,33
CBI – Intelligence	3.755	3.470	3.632	3.922	4.440	3.00+	4,33
CBI – Task Oriented	3.571	3.353	3.298	3.892	3.850	3.02+	4,33

NOTES: Because of small sample sizes, analyses do not include controls for child or program background variables. df = degrees of freedom.

* Significance = p. < .025.

+ Significance = p. < .05.

were consistent over waves, making it difficult to draw any meaningful conclusions from these results. (See Tables A.6.14–A.6.17 in the appendix to this chapter for these results.)

Conclusions

Star ratings were related to a few child outcomes in Wave 1, but these effects did not hold in later waves. Caregiver training and education variables were also associated with some child outcomes, although particular T&E variables were inconsistent across waves. In general, better-educated and more-experienced family child-care providers served children with better cognitive and social outcomes. This finding of some association between the T&E variables and child outcomes is different from what we found for staff in centers, where having a B.A. and more ECE credits were associated with better child outcomes, but years of experience was negatively related to those same outcomes.

We can only speculate about the reason for the reverse effect of years of experience in family child-care homes. Turnover among staff in centers is generally high, not surprising given low starting salaries and limited raises (Helburn et al., 1995). Thus, external incentives for staying are limited. However, in child-care homes, providers are also business owners and entrepreneurs. Those who run child-care enterprises in their homes may do so because the business works well for them by providing some combination of security, a sense of accomplishment, and, for about one-third, an opportunity to earn money while caring for their own children at home (e.g., Kontos, Howes, and Galinsky, 1996; Layzer and Goodson, 2006). While there is certainly turnover, much of it occurs because young mothers who chose to care for others' children in order to stay home with their own children close their doors when their children reach school-age (Zellman et al., 1992). Those who stay in business want to do so. They are not running their business in most cases for want of other options (Helburn, Morris, and Modigliani, 2002).

Findings for the process measures were very limited and inconsistent, suggesting no meaningful relationships between the process measures and either the star ratings or the child outcomes. The general lack of findings here mirrors what we found for the Q-QRIS variables.

These results, based on simple correlations and analyses of variance, do not control for key variables, such as family demographics. Furthermore, the small sample size and inability to detect effects reduce our ability to draw conclusions based on these data. We did find some relationships between star ratings and individual Q-QRIS components and child functioning. But the lack of patterns, inconsistency in the results over waves, and the limited associations between the process measures and child outcomes underscore the many things we do not understand about quality in child-care homes.

CHAPTER SEVEN

Discussion

Key Research Questions

The overarching goal of this study was to examine the validity of the Qualistar QRIS as an indicator of child-care quality and as a policy tool to improve quality. However, we needed to consider measurement of the components of the Q-QRIS before we could study the system as a whole. We assessed the validity of the Q-QRIS by addressing 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, the individual Q-QRIS components, and concurrent child outcomes? Is provider quality related to future child outcomes?
5. 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?
6. Are there subgroups of children for whom the link between measures of child-care quality and child outcomes are stronger?
7. Did center quality change over time? Did child-care center home quality improve over time?

Below, we review our findings on the key questions, discuss the study's limitations, and contextualize the results. We conclude with some implications for the Qualistar QRIS and child-care quality rating systems more generally.

1. What are the characteristics of the Q-QRIS components as measures?

Our early analyses identified significant measurement issues with several Q-QRIS components. As a result, substantial effort was expended on improving the measures of the different components that make up the Q-QRIS. Effort was made to develop a reliable and feasible measure of ratios in this high-stakes context. Over the course of the evaluation, observations replaced self-reports as the data-collection method; the distribution of these observations was informed by our analyses of sign-in/sign-out data. The measure of parent involvement required significant attention and revision over the course of the study as well. Results indicate that the revised measure, the Family Partnership survey of parents and providers, is associated with a number of key outcomes but is unrelated to the other measures of process quality. These find-

ings provide some support for our decision to sacrifice longitudinal data on FP to improve the measure over time; Qualistar now has a better measure that differentiates among programs. However, given that various measures of the FP are related to child outcomes but unrelated to other measures of quality, it is clear that the FP still requires further analysis, including its relationship to other measures that purport to assess similar concepts. Teacher training and education measures still need a good deal of attention as well; we will be focusing on the assessment of movement of staff over the course of the day in future work, as we believe it is fundamental to assessing and understanding the importance of staff training and education. Finally, Qualistar needs to decide whether the cost and effort required for providers to earn NAEYC accreditation is justified given the limited relationships we found between accreditation status and other measures.

2. **How closely related are the five Q-QRIS component measures?**

The component measures correlated moderately well across all waves. Lower ratios and head teacher and director education and experience are associated with higher-quality classroom environments. Accreditation is also associated with higher ECERS-R scores. FP parent and provider points are positively associated with director education level; FP parent points are positively associated with head teacher ECE credits and negatively with ratios, but only in certain waves.

3. **Do providers that receive high scores on the star ratings and the individual Q-QRIS components also receive high scores on process-quality measures?**

We also examined the relationships between the star ratings and the individual Q-QRIS component measures and the process-quality measures that were chosen as criterion measures (the Caregiver Interaction Scale [Arnett, 1989] and the Pre-Kindergarten Snapshot [Howes, 1997]). At Wave 1, providers that scored high on the Q-QRIS components and star ratings scored high on a few of the CIS subscales. Virtually all of these relationships were in the expected direction. However, there were no significant associations between Q-QRIS components and the Pre-K. Not surprisingly, those components that focused more directly on process quality were more closely associated with the process-quality measures. Those that were most structural, e.g., ratios, demonstrated the fewest relationships. However, these findings were not replicated across Waves 2 and 3. Lack of consistency in findings across waves may be due, at least in part, to nonrandom drop-out of providers from the study sample. Taken together, the results suggest that the star ratings and the Q-QRIS components are generally unrelated to the process measures of quality.

4. **Is there a relationship between the Q-QRIS components and concurrent child outcomes? Is provider quality related to future child outcomes?**

We found few relationships between individual Q-QRIS components and child outcomes, some of which were counterintuitive, and virtually none between star ratings and child outcomes. As with the process-quality correlations, the results were not replicated across waves, making interpretation problematic. In Table 7.1, the inconsistency of the findings across waves is apparent; results are clearly wave-specific.

Table 7.1
Relationship of Q-QRIS Components to Child Outcomes: Summary of Cross-Sectional Results, by Wave

Q-QRIS Components	Wave 1	Wave 2	Wave 3
Mean ratio	None	Creativity (CBI)* Dependence (CBI)*	None
ECERS-R	None	None	Distractibility (CBI)*
FP parent survey	None	None	Hostility (CBI)*
FP provider survey	None	None	Independence (CBI)*
Accreditation status	None	None	None
Director B.A.	None	None	None
Director administration experience	None	Considerateness (CBI)* Hostility (CBI)*	None
Head teacher B.A.	None	None	None
Head teacher ECE credits	None	None	None
Head teacher teaching experience	None	None	Apathy (CBI)* Dependence (CBI)* Considerateness (CBI)*

NOTE: All analyses were conducted using regression models.
* Significance = p. < .025.

We found a very similar pattern of results for the cross-sectional analyses that used the process-quality measures (i.e., the Caregiver Interaction Scale and the Pre-Kindergarten Snapshot) to predict child outcomes. The four subscales for each of these two measures did not predict any of the child outcomes.

The family child-care data, based on small numbers, revealed a few relationships between the star ratings and the individual Q-QRIS components and both the process-quality measures and child outcomes. However, here, too, the effects were inconsistent over waves.

5. 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?

Given the lack of associations between the Q-QRIS components and child outcomes, we were unable to address issues of combining or weighting of the Q-QRIS components. The combining of components that addressed quite different aspects of quality was a key aspect of the Qualistar QRIS, and we hoped to be able to provide empirical guidance on how to do it. However, because we did not find large or consistent effects in the analyses of the individual Q-QRIS components, we concluded that it was not meaningful to conduct those analyses.

6. Are there subgroups of children for whom the link between measures of child-care quality and child outcomes are stronger?

We conducted a series of regressions using the covariates and outcomes, described in Chapter Five, with different subgroups of children who came from low-income homes or who had high doses of child-care exposure. The pattern of results for these children did not differ from those found for the general population.

7. Did center quality change over time? Did family child-care quality improve over time? If so, did the Q-QRIS components also improve over time?

Provider quality improved over time, as evidenced by improvement in Environment Rating Scale scores. Another indicator of improvement was the increase in the percentage of accredited centers over the course of the study. Family child-care provider quality also improved slightly. It is likely that the improvement came about through HB 1297 and the Qualistar quality-improvement process. However, it is possible that improvements were a reaction to simply being assessed; it would be extremely useful to know if assessment alone reliably has this effect. Alternatively, improvements reflected regular practice in a group of providers that self-selected into a quality-improvement study. Intervention participant self-selection, the lack of a comparison group, and limited data on the implementation of the intervention made testing the effect of the intervention impossible.

Summary

We assessed the validity of the Q-QRIS by addressing the seven questions above. In doing so, we assessed how well the individual components of the Q-QRIS, as well as the Q-QRIS as a whole, function in a context in which considerable stakes may be attached to a particular rating. As discussed in Chapter One, validation is a complex process that rarely produces a "yes" or "no" answer. In this study, definitive conclusions about the validity of the Q-QRIS and its components are particularly difficult to draw because of the study design. For example, the process measures that we used as our criterion measures were collected from a single classroom in each center, and therefore provide limited information with which we can validate the Q-QRIS and its components as measures of provider quality. Similarly, we attempted to examine the internal structure of the ECERS-R, but because our data were drawn primarily from settings that did not have stakes attached to the ECERS-R rating, it is unclear whether similar findings would be replicated in our current context, where ratings are associated with potential consequences. Likewise, we were able to create a measure of parent involvement that showed variation across programs of varying quality, but little is known about its internal consistency and other psychometric properties. In light of these limitations, our discussion about the validity evidence for the Q-QRIS and its components should not be seen as definitive.

Taken together, our findings provide mixed support for the Q-QRIS and its components as measures of provider quality. The Q-QRIS and the component measures correlate moderately with each other and show some relationships with one of the process measures chosen as criterion measures (CIS). However, the relationships were weak and inconsistent across waves, and the Q-QRIS and its components showed virtually no relationships to the other process measure used as a criterion (Pre-K Snapshot).

Although the logic model described in Chapter One suggests that the Q-QRIS and its components should predict child outcomes, there is little evidence in this study supporting these relationships. The lack of a randomized design coupled with nonrandom provider attrition and very high child attrition in our sample make it difficult for us to generalize our findings to the functioning of similarly constructed QRISs in other settings.

While our study does not allow us to make strong statements about the validity of the Q-QRIS, it does underscore the enormous need for more thoughtful, iterative research on QRISs. Much has been learned from this study about how to design and validate a QRIS. In particular, it is clear that the first research focus should be on the integrity and validity of the

individual QRIS components. Once they have been validated, attention must be focused on how they should be combined to produce a valid measure of overall program quality. Additional research should focus on the relative merits of a range of quality-improvement approaches (e.g. feedback alone, type of coaching). Attention should also be directed to testing each aspect of the logic model underlying QRISs. All of this work will help to build a much richer database on how best to build, implement, and validate QRISs. Given the resources increasingly directed to these systems and the high stakes attached, such work is critical if we are to ensure that providers, children, and families benefit as much as possible from these efforts.

Contextualizing the Child Outcome Findings

In order to understand how our child outcomes findings compare with those of other research studies, we conducted a targeted literature review, looking only for papers that would be directly comparable to the current study.

Papers were identified using the Early Childhood Research and Practice (ECRP) literature reviews and our own searches using the American Psychological Association PsycINFO Database, the Education Resources Information Center Database, and the Economic Literature Database. The database searches focused on papers published in 2001 through 2007 and augmented the papers identified by the ECRP reviews. In total, 35 peer-reviewed papers that linked various indicators of child-care quality to child functioning or child outcomes were identified for review. Of these 35 papers, only papers that examined individual QRIS components—the ECERS-R, child-staff ratios, staff training and education, NAEYC accreditation, or a measure of parent involvement or family partnership—were selected.[1] Studies had to link the quality measures to the functioning or well-being of three- to five-year-old children to be included. These selection criteria left us with 15 papers. Family income in the selected studies varied widely, ranging from quite affluent to poor. This selection process allows us to make direct comparisons between the results reported in these studies and findings from this study. However, it meant that some of the better-known recent studies were excluded, and the total number of comparison studies is small.

None of the 15 papers found links between child functioning and parent involvement or family provider partnership, accreditation, or director training. For all other quality indicators, findings were mixed. While some studies report significant relationships between QRIS components and child outcomes, some find no relationships at all.[2] In the few studies that found such connections, the relationships were weak. At the very least, it appears that the findings

[1] Some papers included only composite scores of quality ratings; these were excluded because we could not link the reported child outcome measures back to one or more of the specific QRIS components. For example, in one paper, a "Teacher Interaction Scale" was created using a combination of items from the ECERS and the Assessment Profile (McCartney et al., 1997). Papers drawing from the Cost Quality and Child Outcomes in Child Care Centers Study frequently combined the ECERS (Harms and Clifford, 1980), the Caregiver Interaction Scale (Arnett, 1989), the Early Childhood Observation Form (ECOF) Teaching Style (Stipek et al., 1992), and the Adult Involvement Scale (AIS) Teacher Responsiveness Scale (Howes and Stewart, 1987) into a single Classroom Practices Index (Peisner-Feinberg et al., 2001). This criterion resulted in the exclusion of a substantial number of the original group of selected studies.

[2] In order to parallel the current findings, some analyses were excluded from papers that were selected for inclusion in this review. For example, analyses that rely on change scores in the ECERS are not reported. Analyses in the same paper that rely on the absolute value of the ECERS at a point in time were included (e.g., Burchinal et al., 2000).

from the current study are not anomalous. (See Tables A.7.1 and A.7.2 in the appendix to this chapter for a summary table of results and a list of included studies.)

Discussion

Few studies allow for direct comparison with our data. This is not terribly surprising given that there are many ways to measure quality and analyze study findings. As noted above, we were looking for studies that used roughly comparable measures and analytic approaches that allowed straightforward comparisons.

It is important to keep in mind that the incidence of significant relationships we found in our literature review is likely inflated as a result of editors' and reviewers' propensity to reject submissions with nonsignificant findings. Furthermore, many of the studies that were reviewed included only a subset of the measures of quality we were interested in. It is possible that data about these other measures of quality had been collected, but, because they were not significantly related to child outcomes, they were not reported in the published paper. For example, ratios are quite easy to collect in studies for which the ECERS-R and other measures that require that extensive time be spent in the classroom are collected. However, they are frequently not reported; we do not know if their absence reflects a decision not to collect them or a lack of association with child outcomes. Similarly, many of the NICHD ECCRN papers do not include ratios or staff training and education in their analyses of child-care quality. Rather, they define quality only in terms of the Observational Record of the Caregiving Environment (ORCE), a process-quality measure developed specifically for that study.

Conclusions

The mixed results we found in the literature are quite consistent with ours. While the general view is that child-care quality is associated with improved child outcomes, close examination of studies that are most comparable to this study suggests that this logical and appealing assumption may not have empirical support.

It is possible that the lack of relationships between child-care quality measures and child outcomes reflects poor measurement of key components of quality. For example, to our knowledge, our work on measuring child-staff ratios provides a first empirical basis for assessing the validity of procedures to capture such ratios. Thus, estimates of ratios generated in past studies may not have captured the construct in the most optimal and reliable way.

In our analysis, we could not address the magnitude of relationships where they do exist. However, as discussed above, effect sizes appear to be small, even when they are significant. Studies that found both significant and nonsignificant effects were most likely to have income-heterogeneous samples; a lack of links between quality indicators and child outcomes was most common in studies that relied on more affluent samples. This conclusion is consistent with the notion supported by major longitudinal studies noted below that child-care quality is most likely to influence the functioning of less-privileged children.

While it makes sense and holds general appeal that improved quality will translate into improved child outcomes, the many factors that shape children over time may swamp the association, at least in the short term. The major longitudinal studies—the Carolina Abecedarian program (Campbell and Ramey, 1995) and the High/Scope Perry Preschool Project program (Weikart, Bond, and McNeil, 1978)—find child effects many years later when contrasted to no intervention at all (see Ramey and Ramey, 2006; Karoly et al., 1998). But these studies provided intensive interventions to very needy children using stronger methodologies than the

current study allowed. Most child-care settings do not provide a standardized intervention, and evaluations of outcomes do not compare a no-treatment condition against a standardized one. These differences may explain the weaker pattern of findings reported here.

Lessons Learned

As accountability increasingly becomes a driving concept in American education, quality rating systems are proliferating in the child-care arena. But virtually no one is focusing serious effort on how to build good systems. Little attention is being paid to determining which components are most important, how best to measure them, or how to weight and combine the component measures to produce the summary rating that is the key output of most of these systems.

Qualistar Early Learning is to be applauded for understanding that good quality rating systems must be evidence-based, built on careful empirical analyses of component measures, revised as needed to improve those measures, and weighted and combined based on empirical data to produce meaningful, defensible ratings. This study reminds us that building quality rating systems is a challenging task. As more states adopt them, it becomes an increasingly important task as well.

As part of a thorough validation process, it is also important to consider whether a QRIS includes unnecessary components or is missing key quality indicators. It is not clear from our analyses whether it is necessary to retain the NAEYC accreditation component. Accreditation is costly and inaccessible for many programs. As for missing components, we have had many discussions about the possible need to incorporate more process measures, particularly those that capture instruction and relationship quality. One instrument, the Classroom Assessment Scoring System (CLASS) (Pianta, LaParo, and Hamre, 2006), is increasingly being used and may provide an alternative to the ECERS-R, provided that research efforts that are currently underway support existing validity data in the CLASS technical manual (see Pianta, LaParo, and Hamre, 2006).

This study's most significant contributions may be found in the work on the component measures. The focus on conceptualizing and measuring parent involvement deserves particular note, because the concept of family partnerships in child-care settings has been given a lot of generalized support but little empirical attention. The Qualistar parent involvement component in place at the study's inception was of little utility, because the measure employed at that time produced no variation across providers of widely differing quality. The current FP measure has a strong conceptual basis and produces some variability. Our work on ratios was also important. Using data that were complicated to collect and analyze, our efforts and analyses enabled Qualistar to develop a defensible, evidence-based approach to the measurement of child-staff ratios.

This focus on the Q-QRIS components points to an important lesson about quality rating systems: Building a QRIS takes time and probably, for efficiency's sake, should be done incrementally. Each construct to be measured must be clearly articulated, designed, tested, and validated in the context in which it will be used. Once each of the components has been found to be well measured and has been validated, an iterative, evidence-based validation process on the QRIS as a whole can begin.

More research needs to be done to determine the best ways to measure components. A good example is the measurement of staff training and education. Many believe that staff

capacity, and particularly director capacity, is the most important structural quality measure, because it is the quality of staff that most directly affects the experience of children in care. But the literature provides only limited guidance concerning which aspects of staff training and education are most important, from whom this information should be collected, or how it should be combined. A simple question—"Is collecting training and education data from just the Head Teacher enough?"—cannot currently be answered. And yet, it is important to know that answer, as well as the answers to many other related questions. These questions and their answers matter because they will help us measure these important concepts better and may make data collection far more efficient.

Another question concerns nonprogrammatic movement among staff, which providers often use as way to stay within mandated ratios while minimizing staffing costs. How much of it occurs? Is it limited to early morning and late-day periods? Classroom-level measures of staff training and education used in quality rating systems, including Qualistar's, implicitly assume no movement: How correct is that assumption? What can and should we do to correct for movement of staff and children over the course of the day in measuring training and education?

This study found limited associations between Q-QRIS components and the child outcome measures. Perhaps we need to reconsider whether our measures of provider quality are appropriate ones: Several of the Q-QRIS component measures (e.g., staff training and education) only proxy aspects of the environment (e.g., the quality of instruction that staff provide) that are associated with child outcomes). We also should reconsider the child outcomes we employ. Is it reasonable to expect that many of these measures, which we know to be heavily influenced by family and child factors, will be affected by what child-care providers do? Are there better indicators of child functioning on which we can hang a quality rating system? Early childhood educators, researchers, and kindergarten teachers are more interested in children's capacity to regulate their emotions, develop trusting relationships with adults, and approach learning in a motivated, efficacious way than they are in the acquisition of pre-academic skills (National Research Council and Institute of Medicine, 2000a). Might we develop and employ more of these sorts of indicators in our examinations of quality rating systems? Alternatively, it may be appropriate, particularly until we can build a stronger empirical basis for our quality measures, to stay away from longer-term child outcomes entirely, focusing instead on program outputs, such as children's engagement in developmentally appropriate tasks in a safe and supportive environment.

Study findings clearly indicate that there is much work to be done before we can confidently implement quality rating systems at scale. As a starting point, a research base must be established that provides data on how to best measure particular components, which components matter most, and how component scores should be combined and weighted to produce the summary ratings that are the key output of these systems. We are aware that a focus on measurement research will slow the rollout of quality rating systems. But we believe that the effort and the delay will be well worth the wait in the longer term. Good research on component measures, weighting schemes, and appropriate criterion measures will produce better rating systems. These efforts will also establish an ongoing process to continuously improve quality rating systems as we learn more about what really matters for children.

Bibliography

Ainslie, R. C. (1990). Family and center contributions to the adjustment of infants in full-time day care. *New Directions for Child Development*, 49, 39–52.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Arnett, J. (1989). Caregivers in day care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541–552.

Beller, E. K., Stahnke, M., Butz, P., Stahl, W., and Wessels, H. (1996). Two measures of the quality of group care for infants and toddlers. *European Journal of the Psychology of Education*, 11, 151–167.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of Royal Statistical Society*, Series B, 57, 289–300.

Birch, S., and Ladd, G. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology*, 35, 61–79.

——— (1998). Children's interpersonal behaviors and the teacher-child relationship. *Developmental Psychology*, 34, 934–946.

Blau, D. (2000). The production of quality in child-care centers: Another look. *Applied Developmental Science*, 4(3), 136–148.

Block, J. H. (1965). *The Child-Rearing Practices Report.* Berkeley, CA: Institute of Human Development, University of California, Berkeley.

——— (1980). *The Child-Rearing Practices Report (CRPR): A Set of Q Items for the Description of Parental Socialization Attitudes and Values.* Berkeley, CA: Institute of Human Development, University of California, Berkeley.

Blunden, D., Spring, C., and Greenberg, L. M. (1974). Validation of the classroom behavior inventory. *Journal of Consulting and Clinical Psychology*, 42(1), 84–88.

Boushey, H., and Wright, J. (2004). *Working Moms and Child Care.* Washington, DC: Center for Economic and Policy Research.

Broberg, A. G., Hwang, C. P., and Chace, S. V. (1993). Effects of day care on elementary school performance and adjustment. Paper presented at the biennial meetings of the Society for Research in Child Development, New Orleans, LA.

Burchinal, M. R., Cryer, D., Clifford, R. M., and Howes, C. (2002). Caregiving training and classroom quality in child care centers. *Applied Developmental Science*, 6(1), 3–11.

Burchinal, M., Howes, C., and Kontos, S. (2002). Structural predictors of child care quality in child care homes. *Early Childhood Research Quarterly*, 17, 87–105.

Burchinal, M. R., Roberts, J. E., Nabors, L. A., and Bryant, D. M. (1996). Quality of center child care and infant cognitive and language development. *Child Development*, 67, 606–620.

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

Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.

Campbell, F. A., and Ramey, C. T. (1995). Cognitive and school outcomes for high-risk African-American students at middle adolescence: Positive effects of early intervention. *American Educational Research Journal*, 32(4), 743–772.

Cassidy, D. J., Buell, M. J., Push-Hoese, S., and Russell, S. (1995). The effects of education on child care teachers' beliefs and classroom quality: Year one evaluation of the TEACH early childhood associate degree scholarship program. *Early Childhood Research Quarterly*, 10, 171–183.

Charlesworth, R., Hart, C. H., Burts, D. C., Thomasson, R. H., Mosley, J., and Fleege, P. O. (1993). Measuring the developmental appropriateness of kindergarten teachers' beliefs and practices. *Early Childhood Research Quarterly*, 8, 255–276.

Chin-Quee, D. S., and Scarr, S. (1994). Lack of early child care effects on school-age children's social competence and academic achievement. *Early Development and Parenting*, 3, 103–112.

Chipty, T. (1995). Economic effects of quality regulations in the day-care industry. *AEA Papers and Proceedings*, 85(2), 419–424.

Clarke-Stewart, K. A., Gruber, C. P., and Fitzgerald, L. M. (1994). *Children at Home and in Day Care*. Hillsdale, NJ.: Earlbaum.

Clarke-Stewart, K. A., Vandell, D. L., Burchinal, M., O'Brien, M., and McCartney, K. (2002). Do regulable features of child-care homes affect children's development? *Early Childhood Research Quarterly*, 17, 52–86.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.

Comer, J. (1984). Home-school relationships as they affect the academic success of children. *Urban Society*, 16, 323–337.

Corbett, H. D., and Wilson, B. L. (1991). *Testing, Reform, and Rebellion.* Norwood, NJ: Ablex Publishing Corporation.

Cost, Quality, and Child Outcomes Study Team (1995). *Cost, Quality, and Child Outcomes in Child Care Centers: Public Report.* Denver, CO: Economics Department, University of Colorado, Denver.

Cryer, D., Hurwitz, S., and Wolery, M. (2000). Continuity of caregiver for infant and toddlers in center-based care: Report on a survey of center practices. *Early Childhood Research Quarterly*, 15(4), 497–514.

Deater-Deckard, K., Pinkerton, R., and Scarr, S. (1996). Child care quality and children's behavioral adjustment: A four-year longitudinal study. *Journal of Child Psychology and Psychiatry*, 37(8), 937–948.

de Kruif, R., McWilliam, R., Ridley, S., and Wakely, M. (2000). Classification of teachers' interaction behaviors in early childhood classrooms. *Early Childhood Research Quarterly*, 15(2), 247–268.

Dekovic, M., Janssen, J., and Gerris, J. (1991). Factor structure and construct validity of the Block child rearing practices report (CRPR). *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3(2), 182–287.

Doherty, G., Forer, B., Lero, D. S., Goelman, H., and LaGrange, A. (2006). Predictors of quality in family child care. *Early Childhood Research Quarterly*, 21, 296–312.

Dry, T., Collins, J., Azer, S., Karolak, E., and Clark, L. (2004). *State Tiered Quality Strategies (TQS), 2004*. Fairfax, VA: National Child Care Information Center.

Duncan, G. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, 74(5), 1454–1475.

Dunn, L. (1993). Proximal and distal features of day care quality and children's development. *Early Childhood Research Quarterly*, 8, 167–192.

Dunn, L. M., and Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Early, D., M. Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., Cai, K., Clifford, R. M., Ebanks, C., Griffin, J. A., Henry, G. T., Howes, C., Iriondo-Perez, J., Jeon, H.-J., Mashburn, A. J., Peisner-Feinberg, E., Pianta, R. C., Vandergrift, N., and Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, 78(2), 558–580.

Elicker, J., Fortner-Wood, C., and Noppe, I. C. (1999). The context of infant attachment in family child care. *Journal of Applied Developmental Psychology*, 20(2), 319–336.

Elicker, J., Noppe, I. C., Noppe, L. D., and Fortner-Wood, C. (1997). The parent-caregiver relationship scale: Rounding out the relationship system in infant child care, *Early Education and Developmen*t, 8(1), 83–100.

Endsley, R. C., and Minish, P. A. (1991). Parent-staff communication in day care centers during morning and afternoon transitions, *Early Childhood Research Quarterly*, 6, 119–135.

Gable, S., and Cole, K. (2000). Parents' child care arrangements and their ecological correlates. *Early Education and Development*, 11(5), 549–572.

Gates, S. M., Zellman, G. L., Moini, J. S., and Suttorp, M. (2006). *Examining Child Care Need Among Military Families*. Santa Monica, CA: RAND Corporation, TR-279-OSD. As of April 28, 2008: http://www.rand.org/pubs/technical_reports/TR279/

Gerber, E. B. (2003). *Predictors of Teacher Sensitive Caregiving in Center-Based Child Care*. Thesis (Ph.D. in Psychology), Berkeley, CA: University of California, Berkeley.

Ghazvini, A., and Readdick, C. (1994). Parent-caregiver communication and quality of care in diverse child care settings. *Early Childhood Research Quarterly*, 9, 207–222.

Goelman, H., Forer, B., Kershaw, P., Doherty, G., Lero, D., and LaGrange, A. (2006). Towards a predictive model of quality in Canadian child care centers. *Early Childhood Research Quarterly*, 21, 280–295.

Goelman, H., and Pence, A. R. (1994). Play, talk literacy, and the ecology of family day care. In H. Goelman and E. V. Jacobs (eds.), *Children's Play in Childcare Settings.* Albany, NY: State University of New York Press, 193–213.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581–586.

Goossens, F. A., Ottenhoff, G., and Koops, W. (1991). Day care and social outcomes in middle childhood: A retrospective study. *Journal of Reproductive and Infant Psychology. Special Issue: International Perspective on Day Care for Young Children*, 9(2–3), 137–150.

Gormley, W. T., Jr., Gayer, T., Phillips, D., and Dawson, B. (2005). The effects of universal pre-K on cognitive development, *Developmental Psychology*, 41(6), 872–884.

Gormley, W. T., Jr., and Weimer, David L. (1999). *Organizational Report Cards.* Cambridge, MA: Harvard University Press.

Hamre, B. K., and Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, 72(2), 625–638.

Harms, T., and Clifford, R. (1980). *Early Childhood Environment Rating Scale.* New York: Teachers College Press.

——— (1989). *The Family Day Care Rating Scale.* New York, NY: Teachers College Press.

Harms, T., Cryer, D., and Clifford, R. (1990). *Infant-Toddler Environment Rating Scale*. New York, NY: Teachers College Press.

——— (1998). *Early Childhood Environment Rating Scale-Revised.* New York, NY: Teachers College Press.

Hart, B., and Risley, T. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*, Baltimore, MD: Brookes.

Hayes, C. D., Palmer, J. L., and Zaslow, M. J. (1990). *Who Cares for America's Children?* Washington, DC: National Academy Press.

Heckman, J. J. (2006). A broader view of what education policy should be. In N. F. Watt, C. Ayoub, R. H. Bradley, J. E. Puma, and W. A. LeBouef (eds.), *The Crisis in Youth Mental Health: Critical Issues and Effective Programs, Volume 4: Early Intervention Programs and Policies*. Westport, CT: Praeger, 3–26.

Helburn, S. W., and Bergmann, B. R. (2002). *America's Child Care Problem: The Way Out*. New York, NY: Palgrave.

Helburn, S. W., Culkin, M. L., Morris, J., et al. (1995). Cost, quality, and child outcomes in child care centers: Public Report. Denver, CO: University of Colorado, Denver.

Helburn, S. W., Morris, J., and Modigliani, K. (2002). Family child care finances and their effect on quality and incentives. *Early Childhood Research Quarterly*, 17, 512–538.

Hofferth, S. L., and Phillips, D. (1987). Child care in the United States, 1970 to 1985. *Journal of Marriage and the Family*, 49, 559–571.

Howes, C. (1988). Relations between early child care and schooling. *Developmental Psychology*, 24, 53–57.

——— (1997). Children's experiences in center-based childcare as a function of teacher background and adult:child ratio. *Merrill-Palmer Quarterly*, 43, 404–425.

——— (2001). Howes Snapshot: Definitions for major sections and categories within sections–revised October 2000. Personal communication.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., and Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*, 23, 27–50.

Howes, C., Hamilton, C. E., and Matheson, C. C. (1994). Children's relationships with peers: Differential associations with aspects of the teacher-child relationship. *Child Development*, 65(1), 253–263.

Howes, C., James, J., and Ritchie, S. (2003). Pathways to effective teaching. *Early Childhood Research Quarterly*, 18, 104–120.

Howes, C., and Matheson, C. C. (1992). Sequences in the development of competent play with peers: Social and social pretend play. *Developmental Psychology*, 28, 961–974.

Howes, C., Matheson, C. C., and Hamilton, C. E. (1994). Maternal, teacher and child care history correlates of children's relationships with peers. *Child Development*, 65(1), 264–273.

Howes, C., and Norris, D. J. (1997). Adding two school-age children: Does it change quality in family child care? *Early Childhood Research Quarterly*, 12, 327–342.

Howes, C., Phillips, D., and Whitebrook, M. (1992). Thresholds of quality: Implications for the social development of children in center-based care. *Child Development*, 63, 449–460.

Howes, C., Rodning, D., Galluzzo, C., and Myers, L. (1988). Attachment and child care: Relationships with mother and caregiver. *Early Childhood Research Quarterly*, 3, 403–416.

Howes, C., and Rubenstein, J. (1981). Toddler peer behavior in two types of day care. *Infant Behavior and Development*, 4, 387–393.

Howes, C., and Smith, E. (1995). Child care quality, teacher behavior, children's play activities, emotional security, and cognitive activity in child care. *Early Childhood Research Quarterly*, 10, 381–404.

Howes, C., Smith, E., and Galinsky, E. (1995). *The Florida Child Care Quality Improvement Study Interim Report.* New York, NY: Families and Work Institute.

Howes, C., and Stewart, P. (1987). Child's play with adults, toys, and peers: An examination of family and child care influences. *Developmental Psychology*, 23, 423–430.

Howes, C., Whitebook, M., and Phillips, D. (1992). Teacher characteristics and effective teaching in child care: Findings from the National Child Care Staffing Study. *Child and Youth Care Forum*, 21(6), 399–414.

Johansen, A. (1992). *Child Care: Preferences, Choice, and Consequences.* Santa Monica, CA: RAND Corporation, N-3237-RGSD. As of April 28, 2008:
http://www.rand.org/pubs/notes/N3237/

Johansen, A., Leibowitz, A., and Waite, L. (1997). *The Importance of Child-Care Characteristics to Choice of Care*. Santa Monica, CA: RAND Corporation, RP-582. As of April 28, 2008:
http://www.rand.org/pubs/reprints/RP582/

Johnson, J. O. (2005). *Who's Minding the Kids? Child Care Arrangements: Winter 2002.* Washington, DC: U.S. Census Bureau. As of April 28, 2008:
http://www.census.gov/prod/2005pubs/p70-101.pdf

Kagan, S. L. (1991). Examining profit and nonprofit child care: An odyssey of quality and auspices. *Journal of Social Issues*, 47(2) (Special Issue on Child Care Policy Research), 87–104.

Karoly, L. A., Greenwood, P. W., Everingham, S. S., Hoube, J., Kilburn, M. R., Rydell, C. P., Sanders, M., and Chiesa, J. (1998). *Investing in Our Children: What We Know and Don't Know About the Costs and Benefits of Early Childhood Interventions*. Santa Monica, CA: RAND Corporation, MR-898-TCWF. As of May 13, 2008:
http://www.rand.org/pubs/monograph_reports/MR898/

Kerbow, D. and Bernhardt, A. (1993). Parental intervention in the school: The context of minority involvement. In: B. Schneider and J. Coleman (eds.), *Parents, Their Children, and Schools.* San Francisco, CA: Westview Press, 115–146.

Kisker, E. E., and Maynard, R. (1991). Quality, cost, and parental choice of child care. In D. M. Blau (ed.), *The Economics of Child Care*. New York, NY: Russell Sage Foundation, 127–143.

Kohn, L., and Zellman, G. L. (1994). *Education in Context: Recent Trends in Child Welfare.* Santa Monica, CA: RAND Corporation, DRU-710-IET. As of April 28, 2008:
http://www.rand.org/pubs/drafts/DRU710/

Kontos, S. (1992). *Family Day Care: Out of the Shadows and into the Light*. Washington, DC: NAEYC.

——— (1994). The ecology of family day care. *Early Childhood Research Quarterly*, 9, 87–110.

Kontos, S., and Dunn, L. (1989). Attitudes of caregivers, maternal experiences with day care and children's development. *Journal of Applied Developmental Psychology*, 10, 37–51.

Kontos, S., Howes, C., and Galinsky, E. (1996). Does training make a difference in quality in family child care? *Early Childhood Research Quarterly*, 11, 427–445.

Kontos, S., Hsu, H.-C., and Dunn, L. (1994). Children's cognitive and social competence in child care centers and family day-care homes. *Journal of Applied Developmental Psychology*, 15(3), 387–411.

Kontos, S., and Wells, W. (1986). Attitudes of caregivers and the day care experience of families. *Early Childhood Research Quarterly*, 1, 47–67.

Lamb, M. E. (1998). Nonparental child care: Context, quality and correlates. In W. Damon, I. E. Siegel, and K. A. Renninger (eds.), *Handbook of Child Psychology, Volume 4: Child Psychology in Practice*. New York, NY: Wiley.

Layzer, J. I., and Goodson, B. D. (2006). *Care in the home: A description of family child care and the experiences of the families and children who use it.* National Study of Child Care for Low-Income Families, Wave I Report. Cambridge, MA: Abt Associates.

Layzer, J. I., Goodson, B. D., and Moss, M. (1993). *Observational study of early childhood programs, Final Report, Volume 1: Life in preschool.* Cambridge, MA: Abt Associates, Inc.

Le, V., Perlman, M., Zellman, G. L., and Hamilton, L. S. (2006). Measuring child-staff ratios in child care centers: Balancing effort and representativeness. *Early Childhood Research Quarterly*, 21, 267–279.

Light, R. J., Singer, J. D., and Willett, J. B. (1990). *By Design: Planning Research on Higher Education*. Cambridge, MA: Harvard University Press.

LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., and Barbarin, O. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly*, 22, 3–17.

Loeb, S., Fuller, B., Kagan, S. L., and Carrol, B. (2004). Child care in poor communities: Early learning effects of type, quality, and stability. *Child Development*, 75(1), 47–65.

Marcon, R. A. (1999). Differential impact of preschool models on development and early learning of inner-city children: A three cohort study. *Developmental Psychology*, 35, 358–375.

McCartney, K., Scarr, S., Rocheleau, A., Phillips, D., Abbott-Shim, M., Eisenberg, M., et al. (1997). Teacher-child interaction and child-care auspices as predictors of social outcomes in infants, toddlers, and preschoolers. *Merrill-Palmer Quarterly*, 43, 426–450.

McNeal, R. B., Jr. (1999). Parental involvement as social capital: differential effectiveness on science achievement, truancy, and dropping out. *Social Forces*, 78(1), 117–144.

Menard, J. (1993). *Development and Implementation of an Integrated Parent Involvement Program for Regular and Special Education Classes in Preschool and Kindergarten for Home and School.* Practicum report. Fort Lauderdale, FL: Nova Southeastern University.

Miedel, W. T., and Reynolds, A. J. (1999). Parent involvement in early intervention of disadvantaged children: Does it matter? *Journal of School Psychology*, 37(4), 379–402.

Mirante, T. J., and Ryckman, D. B. (1974). Classroom behavior inventory: Factor verification. *Journal of Research in Personality*, 8(3), 291–293.

Morris, J. R., and Helburn, S. W. (2000). Child care center quality differences: The role of profit status, client preferences, and trust. *Nonprofit and Voluntary Sector Quarterly*, 29(3), 377–399.

Morrissey, T. W. (2007). *Family Child Care in the United States.* New York, NY: Child Care and Early Education Research Connections. As of April 28, 2008: http://www.researchconnections.org/SendPdf?resourceId=11683

Muller, C. (1993). Parent involvement and academic achievement: An analysis of family resources available to the child. In B. Schneider and J. Coleman (eds.), *Parents, Their Children, and Schools*. San Francisco, CA: Westview Press, 77–113.

Munton, A. G., Rowland, L., Mooney, A., and Lera, M. J. (1997). Using the Early Childhood Environment Rating Scale (ECERS) to evaluate quality of nursery provision in England: Some data concerning reliability. *Educational Research*, 39, 99–104.

NACCRRA—*see* National Association of Child Care Resource and Referral Agencies.

NAEYC—*see* National Association for the Education of Young Children.

National Association for Regulatory Administration (2004). *2004 Family Child Care Licensing Study*.

National Association for the Education of Young Children (no date). NAEYC Academy for Early Childhood Program Accreditation. Web page. As of May 1, 2008: http://www.naeyc.org/academy/standards/

——— (2006). State Policies on Accreditation and Quality Rating System and Tiered Reimbursement Programs. Web page. As of April 28, 2008: http://www.naeyc.org/ece/critical/chart1.asp

National Association of Child Care Resource and Referral Agencies (NACCRRA) (2006a). *2006 Child Care in the State of Michigan*. Arlington, VA: NACCRRA. As of April 28, 2008: http://www.naccrra.org/docs/data/MI.pdf

——— (2006b). *Breaking the Piggy Bank: Parents and the High Price of Child Care.* Arlington, VA: NACCRRA. As of April 28, 2008: http://www.naccrra.org/policy/recent_reports/costreport.php

National Institute of Child Health and Human Development Early Child Care Research Network (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly*, 11, 269–306.

——— (2000). The relation of child care to cognitive and language development. *Child Development*, 71(4), 960–980.

——— (2003). Does quality of child care affect child outcomes at age 4½? *Developmental Psychology*, 39(3), 451–469.

National Research Council and Institute of Medicine (2000a). Committee on Integrating the Science of Early Childhood Development, J. P. Shonkoff, D. A. Phillips, and B. Keilty (eds.), *Early Childhood Intervention: Views from the Field.* Washington, DC: National Academy Press.

——— (2000b). Committee on Integrating the Science of Early Childhood Development, J. P. Shonkoff, and D. A. Phillips (eds.), *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, DC: National Academy Press.

NICHD ECCRN—*see* National Institute of Child Health and Human Development Early Child Care Research Network.

Owen, M. T., and Mulvihill, B. A. (1994). Benefits of a parent education and support program in the first three years. *Family Relations*, 43(2), 206–212.

Owen, M. T., Ware, A. M., and Barfoot, B. (2000). Caregiver-mother partnership behavior and the quality of caregiver–child and mother–child interactions. *Early Childhood Research Quarterly*, 15(3), 413–428.

Paget, K. D. (2001). Review of Early Childhood Environment Rating Scale (Rev. ed.). In B. S. Plake and J. C. Impara (eds.), *14th Mental Measurements Yearbook.* Lincoln, NE: Buros Institute of Mental Measurements.

Peisner-Feinberg, E. S., and Burchinal, M. R. (1997). Relations between preschool children's child-care experiences and concurrent development: The cost, quality, and outcomes study. *Merrill-Palmer Quarterly*, 43(3), 451–477.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan S. L., and Yazejian N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, 72(5), 1534–1553.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M, Yazejian, N., Culkin, M. L., Zelazo, J., Howes, C., Byler, P., Kagan, S. L., and Rustici, J. (1999). *The Children of the Cost, Quality, and Outcomes Study Go to School: Technical Report.* Chapel Hill, NC: University of North Carolina, Frank Porter Graham Child Development Center.

Pence, A. R., and Goelman, H. (1991). The relationship of regulation, training, and motivation to quality of care in family day care. *Child and Youth Care Forum*, 20(2), 83–101.

Perlman, M., Zellman, G., and Le, V. (2004). Examining the psychometric properties of the Early Childhood Environment Rating Scale—Revised. *Early Childhood Research Quarterly*, 19, 398–412.

Phillips, D. A., and Howes, C. (1987). Indicators of quality in child care: Review of research. In D. Phillips (ed.), *Quality in Child Care: What Does Research Tell Us?* Washington, DC: National Association for the Education of Young Children.

Phillips, D. A., Howes, C., and Whitebrook, M. (2002). The social policy context of child care: Effects on quality. In T. A. Revenson, A. R. D'Augelli, S. E. French, D. L. Hughes, and D. Livert (eds.), *A Quarter Century of Community Psychology: Readings from the* American Journal of Community Psychology, New York, NY: Kluwer Academic/Plenum Publishers. xxi, 583, 367–393.

Phillipsen, L., Burchinal, M., Howes, C., and Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly*, 12, 281–303.

Phillipsen, L., Cryer, D., and Howes, C. (1995). Classroom process and classroom structure. In S. W. Helburn (ed.), *Cost, Quality, and Child Outcomes in Child Care Centers*, Denver, CO: Department of Economics, Center for Research in Economics and Social Policy, University of Colorado, Denver, 125–158.

Pianta, R. C. (1994). Patterns of relationships between children and kindergarten teachers. *Journal of School Psychology*, 32, 15–32.

——— (2001). *STRS: Student-Teacher Relationship Scale*. Lutz, FL: Psychological Assessment Resources, Inc.

Pianta, R. C., LaParo, K. M., and Hamre, B. K. (2006). *CLASS—Classroom Assessment Scoring System, Preschool (Pre-K) Version Manual*. Charlottesville, VA: Center for Advanced Study of Teaching and Learning.

Pianta, R. C., Steinberg, M. S., and Rollins, L. (1995). The first two years of school: Teacher-child relationships and deflections in children's classroom adjustment. *Development and Psychopathology*, 7, 297–312.

Powell, D. R. (1978a). Correlates of parent-teacher communication frequency and diversity. *Journal of Educational Research*, 71(6), 333–341.

——— (1978b). The interpersonal relationship between parents and caregivers in day care settings. *American Journal of Orthopsychiatry*, 48(4), 680–689.

Public Law 107-110, The No Child Left Behind Act of 2001, January 8, 2002.

Raikes, H. A., Raikes, H. H., and Wilcox, B. (2005). Regulation, subsidy receipt and provider characteristics: What predicts quality in child care homes? *Early Childhood Research Quarterly*, 20, 164–184.

Ramey, C. T., and Ramey, S. L. (2006). Early learning and school readiness: Can early intervention make a difference? In N. F. Watt, C. Ayoub, R. H. Bradley, J. E. Puma, and W. A. LeBouef (eds.), *The Crisis in Youth Mental Health: Critical Issues and Effective Programs, Volume 4: Early Intervention Programs and Policies*. Westport, CT: Praeger, 291–318.

Reynolds, A. (1992). Comparing measures of parental involvement and their effects on academic achievement. *Early Childhood Research Quarterly, 7,* 441–462.

Rhodes, S., and Hennessy, E. (2000). The effects of specialized training on caregivers and children in early-years settings: An evaluation of the foundations course in playgroup practice. *Early Childhood Research Quarterly*, 15*,* 559–576.

Ritchie, S., and Howes, C. (2003). Program practices, caregiver stability, and child-caregiver relationships. *Applied Developmental Psychology*, 24, 497–516.

Roggman, L. A., Langlois, J. H., Hubbs-Tait, L., and Rieser-Danner, L. A. (1994). Infant day-care, attachment, and the "file drawer problem." *Child Development*, 65(5), 1429–1443.

Rosenthal, R., and Vandell, D. L. (1996). Quality of care at school aged child care programs: Regulatable features, observed experiences, child perspectives, and parent perspectives. *Child Development*, 67(5), 2434–2445.

Roskos, K. and Neuman, S. (1993). Enhancing Head Start parents' conceptions of literacy development and their confidence as literacy teachers: A study of parental involvement. *Early Childhood Development and Care*, 89, 57–73.

Rossbach, H. G. (1990). Assessing the quality of kindergarten environments with the Early Childhood Environment Rating Scale. In J. M. Pieters, K. Breuer, and P. R. J. Simons (eds.), *Learning Environments: Contributions from Dutch and German Research*. New York, NY: Springer Verlag, 77–90.

Ruopp, R., Travers, J., Glantz, F., and Coelen, C. (1979). *Children at the Center: Final Report of the National Day Care Study.* Cambridge, MA: Abt Associates.

Sakai, L. M., Whitebook, M., Wishard, A., and Howes, C. (2003). Evaluating the early childhood ratings scale (ECERS): Assessing differences between the first and revised edition. *Early Childhood Research Quarterly*, 18, 437–445.

Scarr, S. (1998). American child care today. *American Psychologist*, 53(2), 95–108.

Scarr, S., Eisenberg, M., and Deater-Deckard, K. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly*, 9, 131–151.

Schaefer, E. S., Edgerton, M., and Aaronson, M. (1978). *The Classroom Behavior Inventory.* Chapel Hill, NC: University of North Carolina, Chapel Hill.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.

Selvin, M., and Zellman, G. L. (2002). Can the PTA get a passing grade? *Los Angeles Times Magazine*, October 20, 2002, 23–25.

Shepard, L. A., and Dougherty, K. C. (1991). Effects of high stakes testing on instruction. Paper presented at the Annual Meeting of the American Educational Research Association and the National Council on Measurement in Education, New Orleans, LA.

Shimoni, R. (1992). Parent involvement in early childhood education and day care. *Sociological Studies of Child Development*, 5, 73–95.

Shpancer, N. (1997). The link between caregiver-parent relations and children's experiences in daycare and at home: What does the research tell us? *Early Child Development and Care*, 135*,* 7–20.

Shuck, A. (1993). Parents encourage pupils (PEP): An inner-city parent involvement reading project. *The Reading Teacher*, 36(6), 524–528.

Siders, M., and Sledjeski, S. (1978). *How to Grow a Happy Reader: Report on a Study of Parental Involvement as It Relates to a Child's Attitudes and Achievement in the Acquisition of Reading Skills.* Research monograph No. 27. Gainesville, FL, Florida State University.

Stevenson, D., and Baker, D. (1987). The family-school relation and the child's school performance. *Child Development*, 58, 1348–1357.

Stipek, D., Daniels, D., Galuzzo, D., and Milburn, S. (1992). Characterizing early childhood education programs for poor and middle-class children. *Early Childhood Research Quarterly*, 7, 1–19.

Stuart, B., and Pepper, S. (1988). The contribution of caregivers' personality and vocational interests to quality in licensed family child care. *Canadian Journal of Research in Early Childhood Education*, 2(2), 99–109.

Swick, L., and Land, B. (1984). The long-term influence of kindergarten parent involvement program. Paper presented at the Southern Association on Children Under Six Annual Conference, Lexington, KY.

Tietze, W., Bairrao, J., Leal, T. B., and Rossbach, H. G. (1998). Assessing quality characteristics of center-based early childhood environments in Germany and Portugal: A cross-national study. *European Journal of Psychology of Education*, 8, 283–298.

Tonyan, H. A., and Howes, C. (2003). Exploring patterns in time children spend in a variety of child care activities: Associations with environmental quality, ethnicity, and gender. *Early Childhood Research Quarterly*, 18, 121–142.

Tout, K., Zaslow, M., and Berry, D. (2005). Quality and qualifications: Links between professional development and quality in early care and education settings. In M. Zaslow and I. Martinez-Beck (eds.), *Critical Issues in Early Childhood Professional Development*. Baltimore, MD: Paul H. Brookes Publishing Co, 77–110.

U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation (2004). *The Role of Early Head Start Programs in Addressing the Child Care Needs of Low-Income Families with Infants and Toddlers: Influences on Child Care Use and Quality*, Early Head Start Research and Evaluation Project, Washington, DC. As of April 28, 2008: http://www.mathematica-mpr.com/publications/PDFs/roleofearly.pdf

Vandell, D. L., and Wolfe, B. (2000). *Child Care Quality: Does It Matter and Does It Need to Be Improved?* Washington, DC: U.S. Department of Health and Human Services.

Wachs, T. D., Gurkas, P., and Kontos, S. (2004). Predictors of preschool children's compliance behavior in early childhood classroom settings. *Applied Developmental Psychology*, 25, 439–457.

Ware, A. M., Barfoot, B., Rusher, A. S., and Owen, M. T. (1995). The caregiver role in the parent-caregiver partnership: Its relationship to the child care environment. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Indianapolis, IN.

Waters, E. (1990). *Attachment Q Set.* Stony Brook, NY: State University of New York Press.

Weikart, D. P., Bond, J. T., and McNeil, J. T. (1978). *The Ypsilanti Perry Preschool Project: Preschool Years and Longitudinal Results Through Fourth Grade*, Monographs of the High/Scope Educational Research Foundation, Number Three, Ypsilanti, MI: High/Scope Educational Research Foundation.

Whitebook, M. (1996). Accreditation as an indicator of program quality: What research tells us. In S. Bredekamp and B. A. Wiler (eds.), *NAEYC Accreditation: A Decade of Learning and the Years Ahead.* Washington, DC: NAEYC, 31–46.

Whitebook, M., Howes, C., and Phillips, D. (1998). *Worthy Work, Unlivable Wages: The National Child Care Staffing Study, 1988–1997.* Washington, DC: Center for the Child Care Workforce.

Whitebook, M., Phillips, D., Bellm, D., Crowell, N., Almarez, M., and Jo, J. Y. (2004). *Two Years in Early Care and Education: A Community Portrait of Quality and Workforce Stability.* Berkeley, CA: Center for the Study of Child Care Employment, University of California,

Willer, B., Hofferth, S. L., Kisker, E. E., Divine-Hawkins, P., Farqhar, E., and Glantz, F. B. (1991). *The Demand and Supply of Child Care in 1990*. Washington, DC: NAEYC.

Wishard, A. G., Shivers, E. M., Howes, C., and Ritchie, S. (2003). Child care program and teacher practices: Associations with quality and children's experiences. *Early Childhood Research Quarterly*, 18, 65–103.

Woodcock, R. W., and Johnson, M. B. (1990). *Woodcock-Johnson Psycho-Educational Battery—Revised.* Allen, TX: DLM Teaching Resources.

Woodcock, R. W., McGrew, K., Mather, N., and Shrank, F. A. (2006). *Woodcock-Johnson-III Normative Updater*. Chicago, IL: Riverside Publishing.

Zan, B. (2005). NAEYC accreditation and high quality preschool curriculum. *Early Education and Development*, 16(1), 85–102.

Zaslow, M., Halle, T., Martin, L., Cabrera, N., Calkins, J., Pitzer, L., and Margie, N. G. (2006). Child outcome measures in the Study of Child Care Quality. *Evaluation Review*, 30, 577–610.

Zellman, G. L., and Gates, S. M. (2002). *Examining the Cost of Military Child Care*. Santa Monica, CA: RAND Corporation, MR-1415-OSD. As of April 28, 2008: http://www.rand.org/pubs/monograph_reports/MR1415/

Zellman, G. L., and Johansen, A. S. (1996). Investment or overkill: Should military child care centers be accredited? *Armed Forces and Society*, 23, 249–268.

Zellman, G. L., Johansen, A., Meredith, L. S., and Selvin, M. (1992). *Improving the Delivery of Military Child Care: An Analysis of Current Operations and New Approaches*. Santa Monica, CA: RAND Corporation, R-4145-FMP. As of April 28, 2008: http://www.rand.org/pubs/reports/R4145/

Zellman, G. L., and Perlman, M. (forthcoming). Child-Care Quality Rating and Improvement Systems in Five Pioneer States.

Zellman, G. L., and Perlman, M. (2006). Parent involvement in child care settings: Conceptual and measurement issues. *Early Child Development and Care*, 176(5), 521–538.