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

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.