

# EVALUATION OF QUALITY RATING AND IMPROVEMENT SYSTEMS FOR EARLY CHILDHOOD PROGRAMS AND SCHOOL-AGE CARE: MEASURING CHILDREN'S DEVELOPMENT



Research-to-Policy, Research-to-Practice Brief OPRE 2011-11c

June 2011



**DISCLAIMER:**

The views expressed in this publication do not necessarily represent the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families or the U.S. Department of Health and Human Services.

**ACKNOWLEDGMENTS**

The authors would like to thank Ivelisse Martinez-Beck, Mary Bruce Webb and Naomi Goldstein at the Office of Planning, Research and Evaluation and Kathryn Tout at Child Trends for their guidance and feedback on this paper.

# Evaluation of Quality Rating and Improvement Systems for Early Childhood Programs and School-Age Care: Measuring Children's Development

Research-to-Policy, Research-to-Practice Brief OPRE 2011-11c  
June 2011

**Submitted to:**

Ivelisse Martinez-Beck, PhD., Project Officer  
Office of Planning, Research and Evaluation  
Administration for Children and Families  
U.S. Department of Health and Human Services

**Submitted by:**

James Elicker, Purdue University  
Kathy R. Thornburg, University of Missouri

**Contract Number: GS10F0030R**

Project Director: Kathryn Tout  
Child Trends  
4301 Connecticut Ave NW  
Washington DC, 20008

**Suggested Citation:**

Elicker, J. & Thornburg, K. R. (2011). *Evaluation of Quality Rating and Improvement Systems for Early Childhood Programs and School-Age Care: Measuring Children's Development*, Research-to-Policy, Research-to-Practice Brief OPRE 2011-11c. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

This Brief was developed by members of the Quality Initiatives Research and Evaluation Consortium (INQUIRE) which is designed to facilitate the identification of issues and the development and exchange of information and resources related to research and evaluation of quality rating and improvement systems (QRIS) and other quality initiatives. INQUIRE is funded by the Office of Planning, Research and Evaluation through the Child Care and Early Education Policy and Research Analysis and Technical Expertise contract with Child Trends.



## Evaluation of Quality Rating and Improvement Systems for Early Childhood Programs and School-Age Care: Measuring Children’s Development

*Quality Rating and Improvement Systems* (QRIS) for early childhood programs and school-age care are being implemented in more than half of the states as a strategy to improve program quality for young children.<sup>1</sup> QRIS provide parents with information to help them select high quality education and care for their children. QRIS give early childhood programs and school-age care providers incentives and resources to improve quality, working through several manageable steps or levels, plus public recognition for their achieved quality levels. An important long-term goal of QRIS is to improve child development and school readiness outcomes for children. This Brief examines key issues in measuring children’s development when evaluating QRIS.<sup>2</sup>

### Why Assess Child Development in QRIS Evaluations?

Typically, QRIS have five components: a set of quality standards and levels; a rating and monitoring system to assign quality levels; supports for providers to improve quality; incentives to promote providers’ participation; and information for parents and the general public about the significance of quality for children’s development and how QRIS can help them select quality programs.<sup>3</sup> Because improving child outcomes is usually the ultimate goal of QRIS, policy makers and



funders often request or require evidence about linkages between children’s development and the QRIS ratings. For example, funders in Colorado, Indiana, Minnesota, Missouri, Ohio and Tennessee commissioned examinations of children’s development in the ongoing QRIS evaluations in those states.

The findings available to date from these evaluations are mixed, with one study finding significant linkages between QRIS ratings and children’s developmental outcomes and one not finding linkages. In Missouri, children who participated in programs with higher quality ratings and especially low-income children showed significantly higher gains on measures of social-emotional development compared to children in programs with lower ratings.<sup>4</sup> In contrast, in an evaluation of Colorado’s Qualistar program, linkages between the ratings and children’s outcomes were not found.<sup>5</sup> The findings from the other states are not yet available, but will be helpful for building the evidence base on the question of QRIS ratings and linkages with measures of children’s development.

Based on existing theory and research, significant improvements in early childhood program quality should lead to observable gains in child outcomes—especially for low-income children.<sup>6</sup> Therefore the focus of evaluation is often on potential benefits of QRIS to children. This Brief describes the various uses of child development measures in QRIS implementation and outcomes evaluations.<sup>7</sup> Additionally, this Brief outlines the challenges evaluators face when using child development measures to gauge effectiveness of QRIS and offers possible solutions.

## Using Measures of Child Development in QRIS Evaluations

Measures of children’s development may be used in QRIS evaluations for a variety of purposes. For example, child measures may be included to provide:

- **Descriptive information** about the children enrolled in the QRIS. Especially in the early stages of QRIS implementation, program leaders will want information about the types of providers that are participating and the characteristics of the children they serve. This information will help planners determine if the QRIS is reaching all children in targeted populations—or if additional outreach strategies are needed to increase participation among some groups. For example, is the QRIS reaching children with special developmental needs? Are infants and toddlers well represented among participating QRIS providers? Are children at highest risk, or are those whose parents use child care subsidies gaining access to programs participating in the QRIS and its highest quality levels? Descriptive information can also help identify the aspects of child development that need the most support. This information can help QRIS target quality improvement services.
- **Correlational evidence for quality rating validity** for the QRIS. Once the QRIS is established, early childhood professionals and policymakers expect providers rated at higher levels to be providing significantly higher quality programming than those rated at lower levels. As a preliminary step, this expectation can be tested if evaluators independently assess quality in a representative sample of QRIS providers at each quality level. (See for example, Langill et al., 2009.<sup>8</sup>) In addition, stakeholders may expect children at higher QRIS quality levels to be showing benefits in terms of learning and development. Early evaluations of a QRIS can assess children’s development in a representative sample of programs participating in the QRIS, at one or multiple points in time, to determine if children in programs rated at the highest levels are showing significantly greater developmental progress compared to children in programs at the lowest rating levels. For example: Do children who are enrolled in programs with higher QRIS ratings show greater gains towards specified learning goals over time? (See Thornburg et al., 2009; Tout et al., 2010.<sup>9</sup>) Because these studies typically use non-experimental designs, it is important to understand their limitations. Because children and parents may differentially choose or have access to different levels of quality, more rigorous evaluations (with experimental designs) are necessary to establish a causal link between program quality and outcomes. QRIS leaders and evaluators should use these results as a guide for planning, but avoid the temptation to jump to conclusions about the early success or failure of the QRIS, based on such non-experimental data collected early in the program’s implementation. These designs are valuable because they demonstrate *whether* and *how* quality ratings (or components of the ratings) are correlated with measures of children’s development. If children in 4-star rated programs looked similar to children in 1-star programs, for example, QRIS designers would be concerned that the rating tool was not adequately identifying sites at the highest quality level. These designs can also help refine quality measures by providing information about which aspects of quality are most strongly correlated with aspects of children’s development.

- **Outcome evaluation.** What is the ultimate short- or long-term impact on children’s development of individual or combined components of the QRIS? Once the QRIS is well-established, program leaders are satisfied it has been fully implemented, and there has been sufficient time for the QRIS to impact components that are directly linked to children’s well-being and development (at least three years, preferably more), evaluators should consider how to measure benefits to children. However before implementing a costly, high-stakes outcome evaluation study, planners should make sure that:
  1. Expected child outcomes have been precisely specified, are measureable, and are tied to specific quality improvement processes (see the next section for more discussion of QRIS logic models);
  2. The QRIS has been fully implemented as planned and is functioning in a smooth and stable manner;
  3. Child outcome research questions are carefully and specifically formulated;
  4. A careful decision has been made about how to sample children in programs participating in the QRIS, and developmental levels or gains of children in the QRIS are compared with appropriate samples of children not receiving QRIS services; and
  5. Child outcome measures have been reviewed to assure that they are aligned with the quality improvement goals and are valid for the particular populations to be included in the study.

## Tips (and Pitfalls to Avoid) for Assessing Child Development in QRIS Evaluations

Once a careful decision has been made to include measures of child development in a QRIS evaluation, it is recommended that QRIS stakeholders and evaluators address the following issues:

- **Articulate the QRIS program theory of change (logic model) by describing key program process components, how those components will affect early childhood program quality, and how aspects of quality will improve specific child outcomes.** QRIS logic models make explicit the program planners’ assumptions and predictions about the program interventions and the outputs and outcomes that will result from these program change processes. (See Tout et al., 2009 for a discussion of QRIS logic models.<sup>10</sup>) Common pitfalls at this stage are that the program theory of change may be implicit only, or inadequately specified, or that assessed child outcomes may be “mismatched” with the specific quality improvement efforts implemented in QRIS program. Early childhood program quality research to date suggests that well-focused, intensive quality interventions, directed at specific aspects of children’s learning environments are most likely to show measureable effects on children’s outcomes.<sup>11</sup> Key questions to answer include: What specific aspects of quality are targeted by the QRIS? And, through what processes are those quality improvements expected to affect children’s development in specific areas?

- Align research questions about children’s development with the evaluation research design (formative or summative).** Formative evaluations are designed to monitor QRIS program implementation, or to provide information that is useful for program improvement. Summative evaluations are designed to rigorously test the effects of QRIS. Some evaluation designs will not provide valid answers to questions about QRIS effects on children’s developmental outcomes (see Zellman et al., 2011 for a discussion of types of QRIS evaluation designs and their purposes).<sup>12</sup> The following research designs allow answers to specific types of evaluation questions about children, but not others. Examples of appropriate evaluation questions and an overview of designs are discussed below and in Table 1.

  - Descriptive/cross sectional**→What are the developmental characteristics of children served in programs at different levels of the QRIS?
  - Comparative**→ Do sub-groups of children participating in QRIS-rated programs differ in their development and learning?
  - Longitudinal/time series**→How does children’s development change over time while participating in QRIS?
  - Experimental or quasi-experimental**→Does the development or school readiness of children in QRIS-rated programs differ from an equivalent or matched group of children who do not participate in QRIS-rated programs?
- Allow the QRIS to mature before assessing its relationship to children’s development.** Child outcomes may be assessed prematurely, before the QRIS is fully implemented with fidelity to planned activities, and planned QRIS outputs attained. It is recommended that the QRIS have a well-developed and specific program logic model, that the QRIS be fully-implemented with stable program participation, and that child outcomes be specified and measureable using valid, appropriate child development instruments.<sup>13</sup> QRIS program planners should seek advice from experienced scientific consultants before making decisions about evaluating child outcomes. Outcome studies are typically high-stakes and high-cost, so planners need to be sure the evaluation is both timely and well-planned.
- Use appropriate child development measures.** Measures may not be appropriate for the targeted area of child development. Measures need to be valid and focused on the particular developmental domains the program is designed to affect. Some child development measures may not be valid, normed, or standardized for the population in the QRIS program. Child measures selected for a child outcome evaluation need to be focused on the specific child outcomes predicted in the QRIS program logic model. Also, QRIS planners need to be assured that each measure selected to assess children’s development is both valid (measuring what is intended) and reliable (provides consistently accurate information over time and by different reporters), has been used in previous research with similar groups of children, and is appropriate as a measure of development in the population(s) of children served within the QRIS.<sup>14</sup>

- **Ensure that data collectors are well trained.** If child development is not assessed reliably in the field, the study may produce weak or no effects. The validity of child development measures depends on the data collectors' proper and consistent use of each measure. QRIS evaluators need to make sure all individuals collecting data are well-trained in the procedures specified for each measure, and that each data collector's reliability is rigorously assessed over time. Evaluators should plan for checks of the data collectors' reliability at regular intervals to ensure integrity of the data throughout the evaluation.
- **Make an explicit decision about how to assess development among children who are dual language learners.** In most QRIS evaluations, the population served includes significant numbers of children who are dual language learners. Program planners and evaluators should make a clear decision about what approach will be used to assess outcomes with these children. Is the goal to assess the child's total language ability? If so, children should be given the option to respond to questions in English or in another language. Is the goal to assess the child's current proficiency in English? In that case, items may emphasize English mastery.

## Summary and Guidelines

The purpose of this Brief is to provide an overview of issues related to the assessment of child development in QRIS evaluations. The following guidelines should be considered when planning and conducting a QRIS evaluation:

1. Take the time to develop a well-specified program logic model when planning a QRIS. Use the logic model to guide program development and to plan evaluation research.
2. Child development may be measured for a variety of purposes in a QRIS evaluation including the provision of descriptive information about how children are doing, validation of the quality measurement strategy used in the QRIS, and analysis of the outcomes of the QRIS.
3. Use only child development or school readiness measures that have norms and solid evidence for reliability and validity with populations similar to those participating in the QRIS.
4. Make sure the QRIS is fully implemented, the rating system has been rigorously validated, and support services are delivered as planned, *before* assessing the impact of the QRIS on child development outcomes. In most programs, this will be a minimum of three years after QRIS implementation.
5. Using the QRIS logic model as a guide, be sure to measure *QRIS outputs* and *early childhood program quality* as mediating variables, in addition to child development outcomes. This will enable evaluators to test the links between specific aspects of the QRIS program and specific results in the children in the system. In addition, collect demographic characteristics of families to help contextualize the results.<sup>15</sup>
6. Use child development measures that are well-matched to the specific objectives for children specified in the QRIS logic model.
7. Make sure the strongest possible designs are implemented if evaluations are high-stakes; experimental designs provide the most confidence if causal attributions are desired.

**Table 1: QRIS Evaluation: Designs That Use Measures of Child Development**

Evaluation Research Design	How Child Development Measures are Used  <i>Research Questions</i>	Purposes & Advantages  (Examples)	Limitations
Descriptive/Cross Sectional	<p>-Child development measures are used to describe children who are participating in QRIS at one point in time.</p> <p>-----</p> <p><b>-What are the developmental characteristics of children being served at different levels of the QRIS?</b></p> <p>(Formative Evaluation)</p>	<p>-Describe characteristics of children who are participating in QRIS.</p> <p>-Determine if children with specific characteristics are being served by higher- or lower-rated providers.</p> <p>-----</p> <p>-Less costly design than measuring children twice; possible to collect data widely on child participants.</p> <p>-Results may be useful to assess QRIS implementation process.</p>	<p>-Impossible to attribute differences in children’s development to QRIS.</p> <p>-Provides program “snapshot” at one point in time, so not possible to understand ongoing trends or changes.</p>
Correlational/ Comparative <sup>16</sup>	<p>-Measures of child development are correlated with QRIS levels, or demographic subgroups of children in QRIS are compared.</p> <p>-Children’s development may be compared with established norms or data from other QRIS programs.</p> <p>-----</p> <p><b>-Do subgroups of children participating in QRIS differ in their development?</b> (Formative Evaluation)</p>	<p>-Examine children’s developmental outcomes in relation to established norms or other localities.</p> <p>-Determine associations between QRIS levels or other quality assessments and children’s development.</p> <p>-----</p> <p>-Less costly design than measuring children twice; possible to collect data on more child participants.</p>	<p>-Not possible to attribute differences in child development to QRIS.</p> <p>-Not possible to disentangle non-QRIS influences on children’s development.</p> <p>-Existing norms or child data from other samples may not be an appropriate comparison to a particular state sample (for example, if state population has a significant population of dual language learners).</p>

Evaluation Research Design	How Child Development Measures are Used <i>Research Questions</i>	Purposes & Advantages (Examples)	Limitations
Longitudinal/Time Series	<p>-Measure children’s development at two or more points in time to assess change.</p> <p>-----</p> <p><b>-Do children in higher quality programs (more stars) make greater gains than children in lower quality programs?</b> (Formative Evaluation)</p>	<p>-Document rate of change in children’s development based on quality level.</p> <p>-Possible to compare children’s development and rates of change with established norms.</p>	<p>-Does not provide solid evidence that observed changes are due to the QRIS.</p> <p>-More costly than other designs, due to data collection at multiple time points.</p>
Experimental or Quasi-Experimental <sup>17</sup>	<p>-Compare child development outcomes in QRIS to equivalent or matched groups of children who do not participate in QRIS.</p> <p>-----</p> <p><b>-Does the development or school readiness of children in QRIS differ from a matched group of children whose early childhood programs did not participate in QRIS?</b></p> <p><b>-Are children’s gain scores larger when they are in programs rated at higher QRIS quality levels, compared with children at lower levels or with children not in QRIS?</b></p> <p>(Summative Evaluation)</p>	<p>-Provides strongest evidence that change in children’s development is attributed to QRIS.</p>	<p>-May not be practical or ethical to randomly select equivalent experimental and control groups.</p> <p>-May be practical only on a smaller scale, limiting generalizability of results.</p> <p>-Non-randomly selected comparison groups may not be equivalent, threatening validity.</p> <p>-Costly, especially if undertaken on large scale.</p>

**Note:** Research design features described above may be combined for additional advantages.

## Endnotes

- <sup>1</sup> Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2010, April). *Compendium of Quality Rating Systems and Evaluations*, OPRE Report. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- <sup>2</sup> The majority of QRIS include a variety of programs including licensed child care centers, licensed family child care providers, state pre-kindergarten programs, Head Start/Early Head Start, and school-age programs (see endnote i). However, this Brief will focus primarily on issues related to measuring child outcomes for preschool-aged children attending licensed child care centers and family child care programs prior to kindergarten.
- <sup>3</sup> Tout, K., Zaslow, M., Halle, T., & Forry, N. (2009). *Issues for the Next Decade of Quality Rating and Improvement Systems*. OPRE Issue Brief #3. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- <sup>4</sup> Thornburg, K., Mayfield, W. A., Hawks, J. S., & Fuger, K. L. (2009). The Missouri Quality Rating System School Readiness Study. University of Missouri--Columbia. Center for Family Policy and Research. (<http://mucenter.missouri.edu/MOQRSreport.pdf>)
- <sup>5</sup> Zellman, G. L., Perlman, M., Le, V., & Setodji, C. M. (2008). *Assessing the validity of the Qualistar Early Learning quality rating and improvement system as a tool for improving child-care quality*. (MG-650-QEL). Santa Monica, CA: RAND Corporation.
- <sup>6</sup> NICHD Early Child Care Research Network. (2002). Child care structure, process, outcome: Direct and indirect effects of child care quality on young children's development. *Psychological Science*, 13(2), 199-206.
- <sup>7</sup> For a more extensive discussion of QRIS evaluation, see Zellman, G. L., Brandon, R. N., Boller, K., & Kreader, J. L. (2011). *Effective Evaluation of Quality Rating and Improvement Systems for Early Care and Education and School-Age Care*, Research-to-Policy, Research-to-Practice Brief, OPRE 2011-11a. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- <sup>8</sup> Langill, C., Elicker, J., Ruprecht, K., Kwon, K., & Guenin, J. (2009). *Paths to QUALITY- A child care quality rating and improvement system for Indiana: Evaluation methods and measures*. (Technical Report No. 2). West Lafayette, Indiana: Center for Families, Purdue University. ([www.cfs.purdue.edu/cff/publications/publications.html](http://www.cfs.purdue.edu/cff/publications/publications.html))
- <sup>9</sup> Thornburg, K., Mayfield, W. A., Hawks, J. S., & Fuger, K. L. (2009). The Missouri Quality Rating System School Readiness Study. University of Missouri—Columbia. Center for Family Policy and Research. (<http://mucenter.missouri.edu/MOQRSreport.pdf>); Tout, K., Starr, R., Isner, T., Soli, M., Cleveland, J., & Quinn, K. (2010). *Evaluation of Parent Aware: Minnesota's Quality Rating and Improvement System Pilot, Year 3 Evaluation Report*. St. Paul: Minnesota Early Learning Foundation.
- <sup>10</sup> Tout et al, 2009.
- <sup>11</sup> Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict to child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout & T. Halle (Eds). *Quality Measurement in Early Childhood Settings*. Baltimore: Paul H. Brookes Publishing Co.

- <sup>12</sup> Zellman, G. L., Brandon, R. N., Boller, K., & Kreader, J. L. (2011). *Effective Evaluation of Quality Rating and Improvement Systems for Early Care and Education and School-Age Care*, Research-to-Policy, Research-to-Practice Brief, OPRE 2011-11a. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- <sup>13</sup> See resource list in Snow, C. E., & Van Hemel, S.B. (Eds.) (2008). *Early Childhood Assessment: Why, What and How*. Committee on Developmental Outcomes and Assessments for Young Children; Board on Children, Youth, and Families; Board on Testing and Assessment; Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- <sup>14</sup> Zaslow, M., Halle, T., Martin, L., Cabrera, N. J., Calkins, J., Pitzer, L., & Margie, N. G. (2006). Child outcome measures in the study of child care quality. *Evaluation Review*, 30(5), 577-610
- <sup>15</sup> Snow, C. E., & Van Hemel, S.B. (Eds.) (2008). *Early Childhood Assessment: Why, What and How*. Committee on Developmental Outcomes and Assessments for Young Children; Board on Children, Youth, and Families; Board on Testing and Assessment; Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- <sup>16</sup> Mertens, D. M. (2005). *Research and evaluation in education and psychology*. Thousand Oaks: Sage Publications
- <sup>17</sup> Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin