Implications of QRIS Design for the Distribution of Program Ratings and Linkages between Ratings and Observed Quality

Implications of QRIS Design for the Distribution of Program Ratings and Linkages between Ratings and Observed Quality

Kathryn Tout, Nina Chien, Laura Rothenberg and Weilin Li, Child Trends

This document was prepared to accompany other resources on evaluation of Quality Rating and Improvement Systems (QRIS) and other quality improvement initiatives developed by the Quality Initiatives Research and Evaluation Consortium (INQUIRE).

Disclaimer: The views expressed in this publication do not necessarily reflect 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.

This report and other reports sponsored by the Office of Planning, Research and Evaluation are available at http://www.acf.hhs.gov/programs/opre/index.html.

Overview

This Brief compares three hypothetical Quality Rating and Improvement Systems (QRIS) that use different rating structures: block, points, and hybrid. Because the quality standards in the hypothetical QRIS are held relatively constant across structures, analyses can be conducted to determine how structure relates to key QRIS outcomes. Three outcomes are examined: the distribution of programs across ratings levels, the linkages of ratings with measures of observed quality, and the scores of individual quality components within each structure.

Findings indicate that the distribution of ratings is significantly related to structure. Whereas fewer than one-fifth of programs achieved a Level 3 or 4 in the block structure, over 70% of programs achieved a Level 3 or 4 in the points and hybrid structures. Rating levels produced by each of the three structures were significantly correlated with observed quality as measured by the Early Childhood Environment Rating Scale - Revised (ECERS-R). However, the points structure was the only structure to produce quality levels in which observed quality was significantly different between each level. The points structure also captured the greatest range of ECERS-R scores with a 1.61 point spread between Level 1 and Level 4 compared to 0.13 and 1.14 point spreads for the block and hybrid structures respectively. Scores across rating levels in the rating structures showed different patterns for specific quality components, with some domains (Health and Safety, Assessment and Accreditation) scoring high regardless of level and structure, others (Family Partnerships) scoring relatively low and others (Teacher Qualifications and Director Qualifications) demonstrating how quality component scores can differ across structures.

The analyses are limited in their application to QRIS because the data were collected with a unique sample and were not collected in the context of a "real" QRIS. Nevertheless, this Brief offers research evidence that can be useful to QRIS administrators as they weigh different design options and their potential consequences.

Implications of QRIS Design for the Distribution of Program Ratings and Linkages Between Ratings and Observed Quality

Background and Purpose

A Quality Rating and Improvement System (QRIS) is a strategy most states and some localities use to improve the quality of early care and education (ECE) that young children receive and to support parent decision making through the provision of program ratings. The majority of states and territories are implementing, designing or piloting a QRIS.[1] Many of these systems were piloted or had planning phases launched within the last 5 years, reflecting a national trend toward increasing systemic ECE quality improvement and accountability efforts. For example, the federal Race to the Top – Early Learning Challenge application in 2011 required applicants to design and implement a tiered QRIS "that is based on consistent and demanding statewide program standards and that establishes meaningful program ratings" and that promotes quality improvement.[2]

A QRIS typically has five components: standards that define ECE program quality, a rating process to measure and designate quality levels, quality improvement supports (such as technical assistance and training) for programs, financial incentives, and dissemination of ratings to parents and consumers.[3] The content, scope and investment in each component vary widely across states however. This variation creates challenges when trying to compare the outcomes of different QRIS strategies across states and to use that information to guide decision-making.

---

[1] According to the QRIS National Learning Network, nearly all states and territories are implementing or designing a QRIS as of August, 2013. http://www.qrisnetwork.org/sites/all/files/maps/QRIS%20Map,%20QRIS%20National%20Learning%20Network,%20www. qrisnetwork.org%20[Revised%20August%202013].pdf

[2] U.S. Department of Education. (2011). Race to the Top – Early Learning Challenge Application for Initial Funding, CFDA Number: 84.412. Washington, DC: U.S. Departments of Education and Health and Human Services.

[3] Mitchell, A. W. (2005). Stair steps to quality: A guide for states and communities developing quality rating systems for early care and education. Alexandria, VA: United Way of America, Success by 6; Zellman, G. L. and Perlman, M. (2008). Child Care Quality Rating Improvement Systems in Five Pioneer States: Implementation issues and lessons learned. Santa Monica, CA: RAND Corporation.

One salient variation across QRIS is the structure used to determine a program's rating level. States use three primary QRIS structures. *Block* structures specify a set of quality standards at each level of quality. Before a program can move up to a higher level of quality, it must meet all of the standards at that level and those at the lower levels. In contrast, a *points* structure assigns points to each quality standard. This structure adds up the points a program receives and assigns a rating based upon defined point ranges for each quality level. A *hybrid* structure is a combination of a block and points structure. The hybrid approaches vary; a typical example uses blocks to define the two lower levels of the system and points to determine the higher levels of the system. A 2010 Compendium of Quality Rating Systems (the Compendium) reported that half (13) of 26 state and local QRIS examined used block structures, five used points structures, six used a hybrid/combination approach and two couldn't be classified in any of the three approaches.[4]

This variation in structural approaches is noteworthy because it may be related to differences in QRIS functioning or effectiveness. Questions that may be asked include: What are the implications of specifying different rules for combining quality standards and assigning quality levels? Do certain strategies work better for creating quality levels that are distinct and meaningful in a QRIS? Answers to these central questions can be useful as states consider the validation of their QRIS.[5] However, it is difficult to disentangle the implications of differences in quality standards, state context and nuances of the rating process from differences related specifically to QRIS structure. In fact, comparing two state QRIS that both use block structures could be inappropriate given the range of possible differences in the rating criteria, process, and context in the states that could also be linked to the outcomes of interest.

To address the challenges that result from state-to-state comparisons of QRIS structures, this Research Brief uses three "hypothetical" QRIS created using existing national data from the Early Childhood Longitudinal Study – Birth Cohort.[6] By using the same quality standards in each QRIS and changing only the structure and rules for combining the standards, it is possible to gain a picture of how different QRIS structures affect outcomes such as the distribution of programs across the rating levels and the degree to which observed quality differs across each level. The goal of this brief is not to identify a QRIS structure that is more effective than other structures but to provide a descriptive portrait of how decisions about structure are related to program ratings.

Method

The approach used in this Brief is to create hypothetical QRIS that are modeled after existing state QRIS but can be easily manipulated to test the implications of using different QRIS rating structures. These simulated QRIS are developed using data from early care and education programs included in the Early Childhood Longitudinal Study – Birth cohort (ECLS-B). The ECLS-B is a nationally representative sample of children born in 2001 who were followed from 9 months through kindergarten.[7] The current study uses data from the preschool (48 month) wave of data collection.

---

[4] Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., and Boller, K. (2010). Compendium of Quality Rating Systems and Evaluations. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

[5] For further details about QRIS validation, see work supported by the Office of Planning, Research and Evaluation (OPRE) available at: http://www.researchconnections.org/content/childcare/federal/inquire-products.html.

[6] Snow, K., Thalji, L., Derecho, A., Wheeless, S., Lennon, J., Kinsey, S., Rogers, J., Raspa, M., and Park, J. (2007). *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), Preschool Year Data File User's Manual (2005-06) (NCES 2008-024)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

[7] Snow et al., 2007

The full ECLS-B sample contains approximately[8] 10,700 children. The study described in this Brief uses data from the Child Care Observation (CCO) subsample (approximately 1,750 children) which contains data on observed classroom quality and other characteristics of the program each child attended.[9] The CCO oversampled for poor and low-income children and Head Start centers. Eligibility criteria for the CCO specified that the child was in that care arrangement for at least 10 hours per week and that the language of care was English or Spanish. The current study focuses only on center-based programs within the CCO (approximately 1,400 programs). Slightly more than seventy percent of programs in this study were non-profit centers and almost sixty percent were Head Start centers. Additionally, almost seventy percent of centers were considered to be in urban areas (as compared to urban cluster or rural areas).

Data for the study described in this Brief are from the following sources: the center director questionnaire, the teacher (care provider) interview, and the child care observation. For each center-based setting, the center director was asked first for general information about the program. Next, the sampled child's primary provider/teacher in the center was interviewed about his or her own background and experiences, the group environment, and the child's experiences. A subsample of these center-based settings was directly observed and rated using the Early Childhood Environment Rating Scale - Revised (ECERS-R).[10] Additionally, information on child-to-caregiver ratios was obtained through repeated counts of children and providers.[11] We derived three of our quality components (director qualifications, family partnerships, and accreditation) from the center director questionnaire, five (health and safety, curriculum, child assessment, teacher qualification, and teacher training) from the teacher interview, and two (environment/ECERS-R and ratio and group size) from the child care observation.

Creating the Quality Components for the Hypothetical QRIS

We used the Compendium to guide selection of quality components to include in the hypothetical QRIS. The categories included: ratio and group size, health and safety, curriculum, child assessment, director qualifications, environment, teacher qualifications, teacher training, family partnerships, and accreditation. We scanned the ECLS-B data sources to determine whether a quality indicator (item) existed that could be included in one of the ten quality categories. The final set of quality components and indicators included in the hypothetical QRIS represent those that were included in at least four states profiled in the Compendium and that were assessed adequately in the ECLS-B. Notable quality components are missing from the final list. For example, program administration and management is a category included widely in QRIS. However, we identified no ECLS-B items that could serve as a proxy for typical quality indicators in this category. Other QRIS quality categories without adequate proxies in the ECLS-B include licensing compliance, cultural and linguistic diversity, provisions for children with special needs and community involvement. Thus, the hypothetical QRIS in this Brief do not reflect the full range of QRIS indicators being used across states. They are also collected using self-report which is different from the verification processes that are used in QRIS.

Ratio and group size. As part of the child care observation, observers recorded the number of children and adults in the setting. Three to six counts were taken at different times during the observation to create an overall session average.

---

[8] All sample size estimates for the ECLS-B are rounded to the nearest 50 per reporting requirements of the National Center for Education Statistics/Institute of Education Sciences.
[9] In the ECLS-B, each child attended a unique program. Programs are not included more than once in the data.
[10] Harms, T., Clifford, R., & Cryer, D. (1998). *The Early Childhood Environment Rating Scale (Revised Edition).* New York: Teachers College Press.
[11] U.S. Department of Education, National Center for Education Statistics. (2009). *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) 9-Month—Kindergarten 2007 Restricted-Use Data File and Electronic Codebook* (CD-ROM). (NCES 2010-010). Washington, DC: Author.

Health and safety. Teachers responded, on a 4-point scale ranging from "always" to "never," to questions assessing whether the program has an operating smoke detector, has a first-aid kit, has the poison control center number and other emergency numbers by the telephone, and has covers on electrical outlets.

Curriculum. Teachers responded (yes/no) to whether they followed a written curriculum, and if so, whether they received training on use of the curricula.

Child assessment. Teachers responded to one item assessing what methods they use for assessments. Response options were classroom observations/work sampling, testing, both, or "something else".

Director qualifications. Program directors provided information on their highest level of education; whether they have a CDA; whether they have another degree in early childhood education; number of relevant college courses taken (relevant areas include the following: early childhood education, elementary education, special education, curriculum development, English as a second language, child development, teaching methods, and program administration/management), and number of years of experience.

Environment. Observers conducted the ECERS-R in center-based classrooms. The ECERS-R Total Score (scores ranging from 1-7) was calculated for each classroom.

Teacher qualifications. Teachers provided information on their highest level of education; whether they have a CDA, and if not, whether they were working on a CDA; whether they have another degree in early childhood education; number of relevant college courses taken (relevant areas include the following: early childhood education, elementary education, special education, curriculum development, English as a second language, child development, teaching methods, and program administration/management), and number of years of experience.

Teacher training. Teachers were asked whether they ever had training for the care of children under 5, whether they received training in the last 12 months, and how many hours of training they received in the last 12 months. Response options for the number of hours of training were "less than 15 hours," "15-23 hours," and "24 or more hours."

Family partnerships. Directors reported whether and how often teachers meet with parents; how often parents receive written letters describing play and learning activities (response options range from "never" to "daily"); and the percent of parents who participate in the following three ways: as volunteers, in a parent council, or by attending special events and activities (response categories range from 0% to 76-100%).

Accreditation. Directors reported whether the center was accredited by any national, state, or local organization.

Construction of the Block, Points, and Hybrid Rating Structures.

The combination of quality indicators into rating structures adhered closely to methods utilized by state QRIS and described in the Compendium. First, the block, points, and hybrid structures themselves (described in Table 1) were structures frequently used by states. Second, we included quality indicators in the structures if they were similar to those used by at least four states in the Compendium, though most quality indicators included were used by 10 or more states.

Table 1. Requirements for the block, points, and hybrid rating structures

Quality Category	Block	Points	Hybrid (Block Level x Weight)
Ratio and Group Size	Level 1: Ratio =1:20; max group size = 32 Level 2: Ratio =1:12; max group size = 24 Level 3: Ratio = 1:10; max group size = 20 Level 4: Ratio = 1:8; max group size = 16	1 point: Ratio =1:20; max group size = 32 3 points: Ratio = 1:12; max group size = 24 5 points: Ratio = 1:10; max group size = 20 7 points: Ratio = 1:8; max group size = 16	weight = 1.5
Health and Safety	Levels 1 - 4 = "Always" or "Most of the time" for all of following: •Operating smoke detector •First aid kid •Poison control number & other emergency numbers by phone •Cover all electrical outlets	2 points = "Always" or "Most of the time" for all of following: •Operating smoke detector •First aid kid •Poison control number & other emergency numbers by phone •Cover all electrical outlets	weight = 0.5
Curriculum	Level 1 = No curriculum Levels 2 = Follows a written curriculum Levels 3 & 4 = Follows a written curriculum AND receives training on that curriculum	2 points = Follows a written curriculum 4 points = Follows a written curriculum AND receives training on that curriculum	weight = 1
Environment	ECERS total score: Level 1 = 3.0-3.49 Level 2 = 3.50-4.24 Level 3 = 4.25.-4.99 Level 4 = 5.00-7.00	ECERS total score: 2 points = 3.50-4.24 4 points = 4.25.-4.99 6 points = 5.00-7.00	weight = 1.5
Child Assessment	Levels 1 & 2 = No assessments Levels 3 & 4 = Ratings based on class observations/work OR standardized testing/assessments OR "something else"	3 points = Ratings based on class observations/work OR standardized testing/assessments	weight = 1
Director Qualifications	Level 1: CDA or some college, no degree Level 2: AA or other ECE degree Level 3: AA and 4+ college courses OR AA and CDA/other ECE degree OR Bachelor's *required for all:* 1+ years experience Level 4: Bachelor's and 6+ college courses or Bachelor's and CDA/other ECE *required for all:* 3+ years experience	1 point = CDA or some college, no degree 2 points = AA or other ECE 3 points = Bachelor's 4 points = A graduate degree *********************************** 1 point = 4+ college coursework in ECE, elementary education, special education, curriculum development, ESL, child development, teaching methods, program administration/management 2 points = 6+ college courses *********************************** 1 point = 2-4 years experience 2 points = 5+ years experience	weight = 1.5

Quality Category	Block	Points	Hybrid (Block Level x Weight)
Teacher Qualifications (for focal child's teacher)	Level 1: Some college but no degree; or voc/tech program but no diploma; or voc/tech diploma; or currently working on a CDA Level 2: CDA Level 3: AA and 2+ college courses OR AA and CDA/other ECE degree Level 4: Bachelor's and 6+ college courses or Bachelor's and CDA/other ECE *required for all:* 3+ years experience ********************************* College courses are in the following areas: ECE, elementary education, special education, curriculum development, ESL, child development, teaching methods, program administration/management	2 points = Some college but no degree; or voc/tech program but no diploma; or voc/tech diploma; or currently working on a CDA 4 points = CDA 6 points = AA or other ECE degree 8 points = Bachelor's degree or higher ********************************* 1 point = 4+ college coursework in ECE, elementary education, special education, curriculum development, ESL, child development, teaching methods, program administration/management 2 points = 6+ college courses OR other ECE degree ********************************* 1 point = 3+ years experience	weight = 2.5
Training (for focal child's teacher)	Level 1 = No requirement Level 2 & 3 = 1-23 hours ECE training hours in last 12 months Level 4 = 24+ hours ECE training hours in last 12 months	1 point = Have ever received training in care of children under 5 2 points = 1-23 hours ECE training hours in last 12 months 3 points = 24+ hours ECE training hours in last 12 months	weight = 1.5

Quality Category	Block	Points	Hybrid (Block Level x Weight)

Family Partnerships	Level 1: no requirement
	
	Level 2:
	•Teacher parent meetings = at least once a year
	•Written letters describing play and learning activities (other than written lesson plans)= once a month or more often
	•1 or more parent activities (see below)
	
	Level 3
	•Teacher parent meetings = at least once a year
	•Written letters describing play and learning activities = once a week or more often
	•2 or more parent activities (see below)
	
	Level 4:
	•Teacher parent meetings = at least twice a year
	•Written letters describing play and learning activities = once a week or more often
	•3 parent activities (see below)
	
	*******************************
	
	Parent activities:
	  •>0% of parents volunteer
	  •>0 % parents are members of a parent
	   council
	  •>0 % parents attend special
	   events/activities
		1 point = Teacher parent meetings at least once a year
		2 points = Teacher parent meetings at least twice a year
		
		**********************************
		
		1 point = Written letters describing play and learning activities at least once a **month**
		2 points = Written letters describing play and learning activities at least once a **week**
		
		**********************************
		
		1 point for each of following:
		  •>0% of parents volunteer
		  •>0 % parents are members of a parent council
		  •>0 % parents attend special events/activities
			weight = 0.5

Accreditation	Level 4 = Accredited	3 points = Accreditation	weight = 1

Third, the precise way in which levels and points were assigned to various responses mimicked, to the greatest extent possible, decisions made by states. Because there was variation among states and the strategies used, we aimed to make decisions that represented the middle of the range.

In the **block** structure, all the quality indicators in one level must be met before moving on to the next higher level. In other words, programs received the rating level that is equal to their lowest-rated quality category. Table 1 contains details on how levels were assigned for each quality category. If programs did not meet all required indicators for Level 1, they were assigned a Level 0.

In the **points** structure, we assigned points to programs for each quality indicator and then added them together. We applied cut-offs to the points to create four rating levels: Level 1 - 12 to 23 points; Level 2 - 24 to 32 points; Level 3 - 33 to 40 points; Level 4 - 41 to 48 points (no programs had fewer than 12 points). Table 1 contains details on how points were assigned for various response options for each quality indicator.

In the **hybrid** structure, we calculated ratings levels by multiplying the level achieved for each quality category in the block structure by an assigned weight. We assigned quality categories with a stronger research basis a higher weight than quality indicators with less robust research evidence. For example, if a program achieved a Level 4 for the environment quality category in the block structure, the program would receive 4 points (corresponding to the block level)multiplied by 1.5 (the weight for the environment quality category), or six points in the hybrid structure for the environment quality category.

Table 1 displays the weights that we applied to each quality indicator. After multiplying each quality category level from the block structure by the specified weight, we summed the total points across quality categories in the same way as in the points structure. We then applied cut-offs to the summed total to create four rating levels: Level 1 -18 to 26 points; Level 2 - 27 to 34 points; Level 3 - 35 to 39 points; and Level 4 - 40 to 44 points.[12]

Results

What is the distribution of programs across rating levels in each of the different structures?

The distribution of programs across the quality ratings varied depending on which rating structure was used. Overall, programs received lower ratings in the block structure, higher ratings in the points structure, and middle to high ratings in the hybrid structure. In the block structure (see Figure 1), five percent of programs did not receive a rating because they didn't meet the requirements for Level 1. A large percentage of programs (56%, about 750 programs) were rated at Level 1, with smaller percentages of programs at Levels 2 (22%) and 3 (16%); few programs received a rating at Level 4 (about 50 programs). In contrast, in the points structure (see Figure 1), the majority of programs were rated at Level 3 (45%) and Level 4 (36%). Smaller percentages were rated at Level 2 (16%), and at Level 1 (3%). The hybrid structure (see Figure 1) was more similar to the points structure than the block structure, with higher percentages of programs rated at Levels 2, 3 and 4 (26%, 42%, 28% respectively), and only 4% of programs at Level 1.

---

[12] No programs had fewer than 18 points.

Figure 1. Distribution of ratings across the block, points and hybrid rating structures
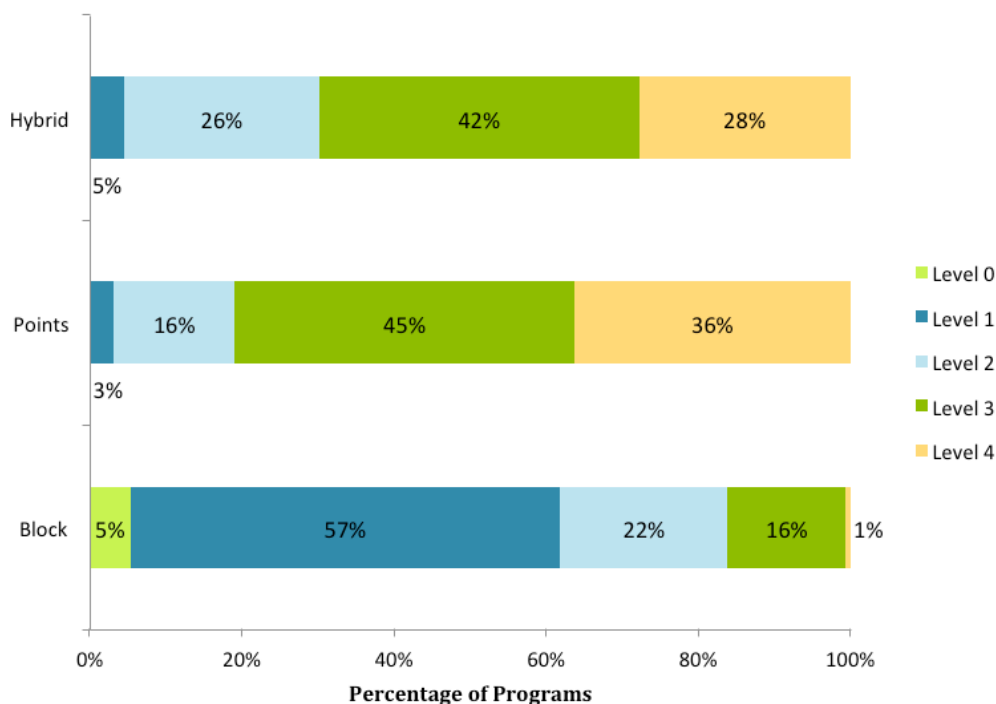
The differences in the distributions of programs produced by each rating structure were statistically significant.[13]

Is observed classroom quality different at the quality levels identified in the rating structures?

Observed classroom quality measured by the ECERS-R was compared across the ratings produced by each rating structure. Because scores on the ECERS-R were used in determining the rating level, we developed new program ratings by removing the ECERS-R scores from the calculations and assigning a new rating. The new calculation resulted in some shifts across how programs were rated, depending on the structure. For the block structure, 76% of programs retained their original rating when ECERS-R scores were removed from the rating criteria. The remainder of programs were either unable to be rated because of missing data in the block structure (19%) or moved down a level (5%). In the points structure, 80% of programs retained their rating, 15% of programs moved up a level, and 5% moved down a level. In the hybrid structure, 65% retained their ratings, 28% moved down a level, and 7% moved up a level. This pattern of movement suggests that the hybrid structure was the most "sensitive" to the inclusion of the ECERS-R scores (relative to the other structures) because the greatest movement in ratings was observed in this structure when ECERS-R scores were no longer included in the rating.

Across each of the rating structures, ECERS-R scores were significantly different across rating levels (see Table 2).[14] At the higher levels of the ratings (between Level 3 and Level 4), the points structure was the only structure that significantly distinguished observed quality (noted as "4 > 3" in Table 2).[15] In addition, the points structure captured the greatest range of ECERS-R scores with a 1.61 point spread between Level 1 and Level 4, compared to a 0.13 and a 1.14 point spread for the block and hybrid structures respectively.

---

[13] The finding was statistically significant > 700; $p < .001$).
[14] The overall F statistic was 22.77 for the block structure, 64.84 for the points structure, and 40.52 for the hybrid structure (all significant at $p < .001$).
[15] This finding was statistically significant at $p < .001$.

Table 2. Mean ECERS-R scores at each rating level, by rating structure, with significant differences noted.

	Level 0	Level 1	Level 2	Level 3	Level 4	Significant Differences between Levels*	Range from Level 1 to 4
Block	3.69	4.41	4.82	4.80	4.54	4>0, 3>1, 3>0, 2>1, 2>0	0.13
Points	.	3.22	3.95	4.49	4.83	4>3, 4>2, 4>1, 3>2, 3>1, 2>1	1.61
Hybrid	.	3.72	4.32	4.70	4.86	4>2, 4>1, 4>0, 3>2, 3>1, 3>0, 2>1, 2>0	1.14

*$p<.05$

Figure 2 provides a graphical depiction of the positive relationship between quality rating level and ECERS-R scores. It is clear that the relationship between rating level and observed quality is slightly stronger using the points and hybrid structures (r = .31 and .34) than using the block structure (r = .20).

Figure 2. ECERS-R scores at each rating level, by rating structure
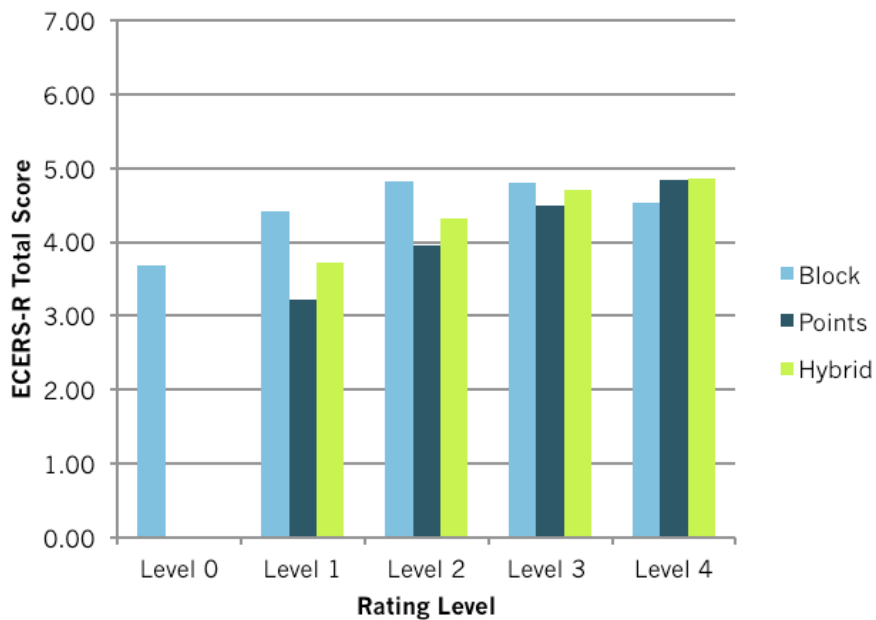
How does scoring of the ten quality components differ across the rating structures?

The rating levels that are produced in each structure are derived from scores on ten different quality components as described in Table 1. Though the criteria for scoring each of the quality components are different across rating structures, the constructs that are measured are theoretically the same. Therefore, it is useful to examine whether and how the individual scores on the quality components differ across the rating levels produced by each of the rating structures (see Table 3). Table 3 displays the average score on the individual quality components across programs rated at each rating level. For example, the average score on the ratio and group size quality category for programs rated at a Level 1 in the block structure was 3.22 as compared with scores of 2.25 and 2.35 for programs rated at a Level 1 in the points and hybrid structures respectively.

Table 3. Average quality component scores at each level, for each rating structure

Quality Components	Ratio and Group Size	Health and Safety	Curriculum	Environment[2]	Assessment	Director Qualifications	Teacher Qualifications[3]	Training	Family Partnerships	Accreditation
Overall[1]	3.35	3.57	3.48	2.82	3.99	3.21	2.99	3.01	2.61	3.53
Block										
Level 0 (n=75)	2.99	3.72	3.63	1.99	3.97	1.38	0.48	2.63	2.19	3.29
Level 1 (n=790)	3.22	3.27	3.19	2.59	3.99	3.09	2.93	2.78	2.44	3.50
Level 2 (n=308)	3.48	4.00	3.82	3.08	3.98	3.61	3.22	3.40	2.57	3.61
Level 3 (n=217)	3.71	4.00	4.00	3.54	4.00	3.82	3.66	3.41	3.66	3.61
Level 4 (n=8)	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
Points										
Level 0 (n=0)										
Level 1 (n=45)	2.25	3.53	3.02	1.40	3.96	1.31	0.56	2.18	1.41	3.09
Level 2 (n=221)	2.80	3.52	3.24	1.90	3.97	2.41	1.75	2.59	2.01	3.25
Level 3 (n=625)	3.30	3.49	3.40	2.70	3.99	3.25	3.04	2.98	2.56	3.47
Level 4 (n=506)	3.74	3.71	3.74	3.51	4.00	3.66	3.67	3.31	3.02	3.77
Hybrid										
Level 0 (n=0)										
Level 1 (n=64)	2.35	3.58	3.14	1.39	3.97	1.31	0.53	2.20	1.65	3.19
Level 2 (n=358)	2.99	3.53	3.36	2.21	3.98	2.51	2.04	2.56	2.28	3.38
Level 3 (n=588)	3.40	3.50	3.39	2.88	4.00	3.45	3.26	3.06	2.71	3.55
Level 4 (n=388)	3.76	3.73	3.81	3.53	3.99	3.86	3.84	3.47	2.95	3.71

[1] Means for the "Overall" row were calculated for children in the observational sample in center-based programs, not for the entire sample in the ECLS-B dataset. These numbers represent means for between 1,011 and 1,420 programs depending on the number of cases for which we have data on each quality category.

[2] Analyses for this characteristic were run using the rating structure that did not include the ECERS as a quality component.

[3] The teacher qualifications category does not capture characteristics of all teachers in a program but instead only captures the focal child's teacher characteristics.

Reviewing the scores within and across structures revealed noteworthy patterns. For example, regardless of the rating structure, scores across the rating levels in the categories of Health and Safety, Assessment and Accreditation are uniformly high (see Figure 3 for the Assessment category example). In contrast, scores in the Family Partnership category are consistently lower across rating levels (see Figure 4). A third pattern is evident in the categories of Director Qualifications and Teacher Qualifications in which scores at all levels of the block structure are high in contrast with the patterns for the points and hybrid structures which increase incrementally by level (see Figure 5 for the Teacher Qualifications category example and notice the particularly wide gap between the scores for each structure at Level 1 and Level 2). This pattern is to be expected based on the scoring rules for points and hybrid structures, but it also highlights how a low overall rating in the block structure can "mask" high scores in individual quality categories.

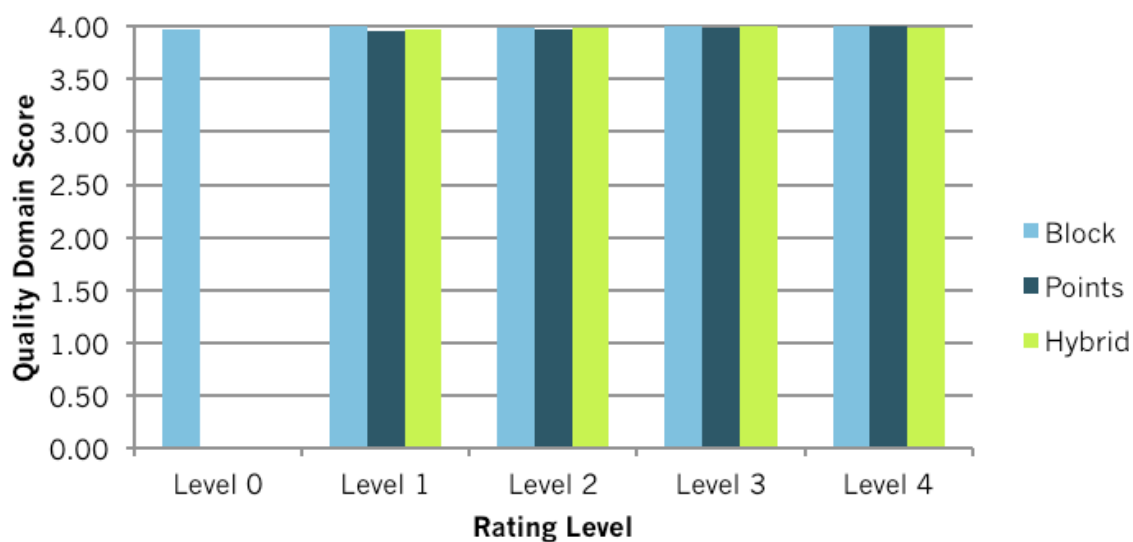*Figure 3.* Scores in the Assessment category, by rating structure and level

Figure 4. Scores in the Family Partnership category, by rating structure and level
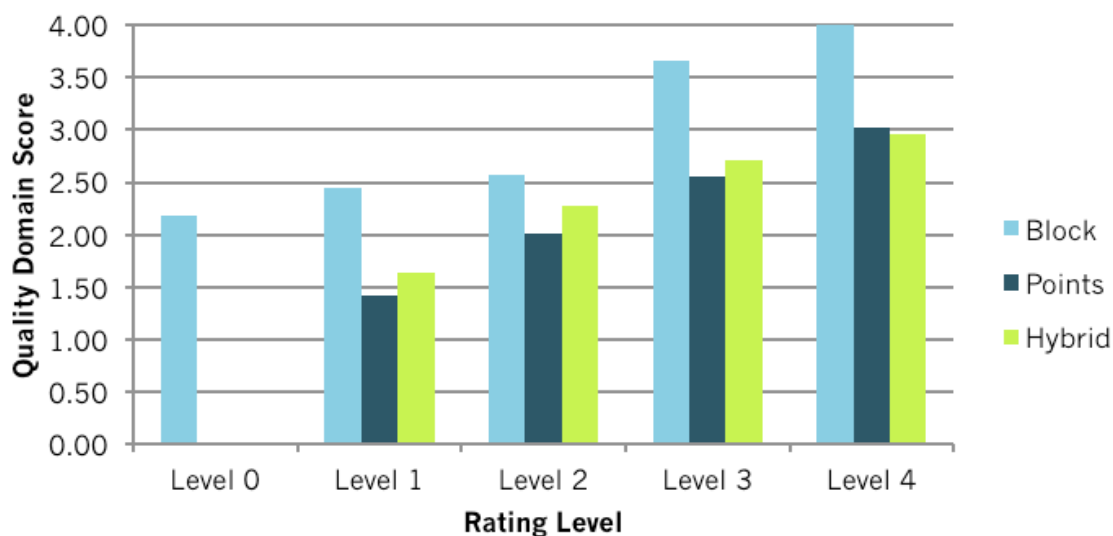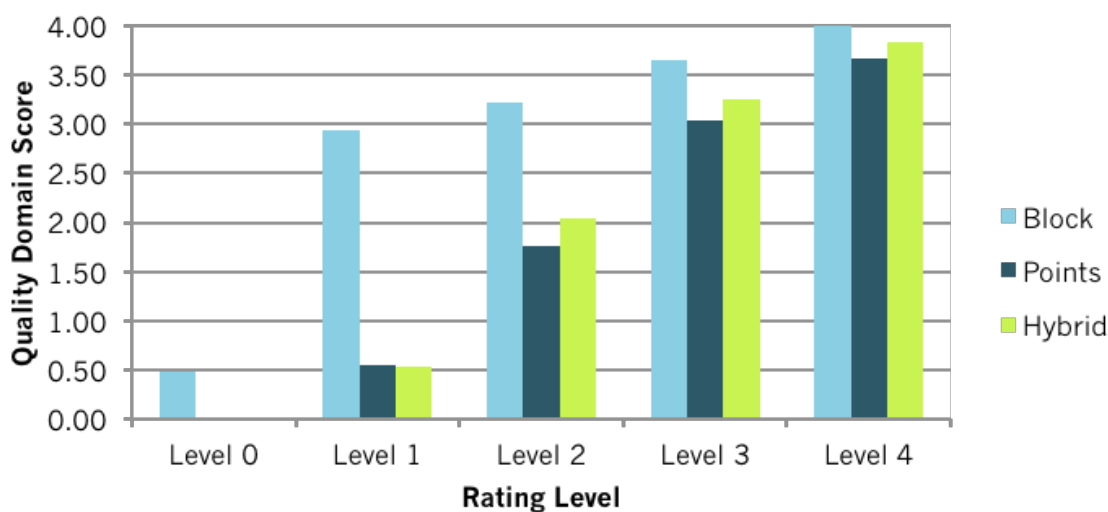
Figure 5. Scores in the Teacher Qualification category, by rating structure and level

Summary and Implications

The purpose of this Brief is to compare three hypothetical QRIS that use different rating structures – block, points, and hybrid. For each structure, we examine the distribution of programs across ratings levels, linkages of ratings with measures of observed quality, and scores on individual quality categories. Overall, the findings in this Brief indicate that QRIS structure has significant implications for these key QRIS outcomes. First, the findings indicate that the distribution of ratings is significantly related to structure. Whereas fewer than one-fifth of programs achieved a Level 3 or 4 in the block structure, over 70% of programs achieved a Level 3 or 4 in the points and hybrid structures.

In addition, rating levels produced by each of the three structures were significantly correlated with observed quality as measured by the ECERS-R. The points structure was the only structure in which observed quality was significantly different between each level. The points structure also captured the greatest range of ECERS-R scores with a 1.61 point spread between Level 1 and Level 4 compared to 0.13 and 1.14 point spreads for the block and hybrid structures respectively.

Finally, scores on individual quality categories showed different patterns for specific quality components in the three structures, with some categories (Health and Safety, Assessment and Accreditation) scoring high regardless of level and structure, others (Family Partnerships) scoring relatively low and still others (Teacher Qualifications and Director Qualifications) demonstrating how quality category scores can differ across structures. In the case of Teacher Qualifications and Director Qualifications, the block structure "masked" higher scores on these categories, such that programs with a low overall rating still had high scores on Teacher and Director Qualifications.

The findings presented have implications for QRIS design decisions and can be used to make predictions about how different structures will influence rating distributions and linkages with observed quality. However, the intent of the analyses in this Brief is not to suggest that one model is better than the others but to inform QRIS design and validation discussions by highlighting the potential role of QRIS structure in outcomes for programs.

The use of secondary data for the analyses presented in the Brief offers important advantages for QRIS research. Using existing data, for example, allowed us to create hypothetical QRIS that could be more easily compared than those created under actual state QRIS criteria. However, the data are limited in their application to QRIS because they were collected for a different purpose with a unique sample. For example, the sample was overrepresented by Head Start programs and child care centers in urban areas. The data were not collected using the same methods as a typical QRIS, across the full range of quality categories,or with the same consequences for programs as in a "real" QRIS. Therefore, we recommend that the findings be approached with these limitations in mind. Nevertheless, this Brief offers research evidence that can be useful to QRIS administrators as they weigh different design options and their potential consequences.