What's Happening

March 2015

Examining changes to Michigan's early childhood quality rating and improvement system (QRIS)

Ann-Marie Faria
Laura E. Hawkinson
Ariela C. Greenberg
Eboni C. Howard
Leah Brown
American Institutes for Research

Key findings

In 2012 Michigan implemented Great Start to Quality, a voluntary quality rating and improvement system (QRIS) that uses a multidimensional assessment to rate the quality of early childhood education programs. Changes to the rating calculation approach announced in 2013 provided an opportunity for Regional Educational Laboratory (REL) Midwest to examine how the changes would affect program quality ratings. Under the revised calculation approach (version 2.0), approximately one-third of programs had a higher self-assessment rating, though the underlying data and measures of program quality were unchanged. A simple alternative total score approach developed by REL Midwest that eliminated criteria for domain scores on the self-assessment yielded rating distributions that were nearly identical to those from the version 2.0 approach. These findings suggest that incremental changes to how QRIS ratings are calculated can alter inferences about program quality.

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

REL MIDWEST
Regional Educational Laboratory
At American Institutes for Research

U.S. Department of Education
Arne Duncan, *Secretary*

Institute of Education Sciences
Sue Betka, *Acting Director*

National Center for Education Evaluation and Regional Assistance
Ruth Curran Neild, *Commissioner*
Joy Lesnick, *Associate Commissioner*
Amy Johnson, *Action Editor*
Chris Boccanfuso, *Project Officer*

REL 2015–029

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

March 2015

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-12-C-0004 by Regional Educational Laboratory Midwest administered by the American Institutes for Research. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Faria, A.-M., Hawkinson, L. E., Greenberg, A. C., Howard, E. C., & Brown, L. (2015). *Examining changes to Michigan's early childhood quality rating and improvement system (QRIS)* (REL 2015–029). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.

Summary

A quality rating and improvement system (QRIS) is a multidimensional assessment system created to rate the quality of early childhood education and child care programs, encourage participating programs to provide higher quality experiences for young learners, and increase the amount of clear and reliable information available to families to help inform their decisions in choosing the best program for their children.

Michigan developed its QRIS, Great Start to Quality, in the early 2000s and rolled it out statewide in 2012. Great Start to Quality's rating calculation approach draws on a self-assessment and an independent observation of quality and rates programs from level 1 (lowest) to level 5 (highest). The original rating approach (version 1.0) used total points and domain scores from the self-assessment. In June 2013 the state changed the approach (version 2.0), reducing the requirements for domain scores on the self-assessment necessary to qualify for the middle QRIS scores (2–4). Recent studies on QRISs have found that the choice of calculation approach can considerably alter the distribution of ratings (Tout, Chien, Rothenberg, & Li, 2014). The 2013 changes to Great Start to Quality provided an opportunity for Regional Educational Laboratory (REL) Midwest to examine how the changes would affect ratings.

This study uses data on more than 11,000 Michigan early childhood education programs (including private center-based programs, Head Start programs, Early Head Start programs, state prekindergarten programs, and family child care providers). It documents how programs were rated under the version 1.0 approach of Great Start to Quality and examines how changes under the version 2.0 approach affected ratings.

Under the version 1.0 approach, almost three-quarters of participating programs that completed a self-assessment rated themselves as being either of lowest or highest quality, and relatively few rated themselves as being of moderate quality. Self-assessment ratings were consistently higher than ratings on the independent observation of quality among the limited number of programs with both types of data. These findings suggest that the self-assessment ratings and ratings on the independent observation may measure different aspects of quality (that is, structural quality for the self-assessment ratings and process quality for the ratings on the independent observation).

Approximately one-third of programs had a higher self-assessment rating under the version 2.0 approach than under the version 1.0 approach, and more programs were rated as being of moderate to high quality, though the underlying data and measures of program quality were unchanged. These findings suggest that incremental changes to how QRIS ratings are calculated can alter inferences about program quality.

REL Midwest developed a simple alternative total score approach that eliminated criteria for domain scores on the self-assessment. This alternative approach resulted in rating distributions that were nearly identical to those under the version 2.0 approach, indicating that the individual domains had limited utility in discriminating overall program quality. However, including domain scores in the calculation of a QRIS is inherently linked to how states define high quality, and removing these requirements may alter the theoretical approach to quality in the QRIS.

This study's findings suggest that Michigan and other states should carefully consider how approaches to calculating QRIS ratings, and changes to these approaches, can affect ratings and resulting inferences about quality of early childhood programs. The study highlights the need for understanding the reliability and validity of instruments used to measure quality and the importance of examining the overall validity of QRISs, considering the high-stakes nature of the ratings.

Contents

Boxes

Figures

Tables

Why this study?

This Regional Educational Laboratory (REL) Midwest study used data shared by Michigan to examine how recent changes to the state's quality rating and improvement system (QRIS) have implications for QRIS administrators, early childhood care and education programs and providers, researchers, and families.

Quality rating and improvement systems are widely used by states to measure and improve quality in early childhood education programs, but there is no consensus on how to best calculate the ratings

As documenting and improving early childhood program quality have become higher national priorities, QRISs—multidimensional assessment systems created to rate the quality of early childhood education and child care programs, encourage early childhood education and child care settings to provide higher quality experiences for young learners, and increase the amount of clear and reliable information available to families about these programs—have expanded rapidly. With attention focused on the potential of high-quality early childhood education to reduce school-readiness gaps (for example, Christenson & Reschly, 2010), all states but Missouri[1] have begun to implement or plan to implement some form of QRIS (QRIS National Learning Network, 2014).

With attention focused on the potential of high-quality early childhood education to reduce school-readiness gaps, all states but Missouri have begun to implement or plan to implement some form of quality rating and improvement system

In the 1990s, QRISs grew out of early tiered reimbursement strategies in the child care subsidy system to promote and reward high-quality care. Thus they initially focused on licensed, subsidized child care centers. In the mid-2000s the focus shifted from using QRISs to rate subsidized programs to using QRISs as a policy mechanism to improve early child care quality across multiple program settings. States have expanded their QRISs to include a diverse range of program types (for example, school-based prekindergarten programs, state-funded prekindergarten programs, Head Start and Early Head Start programs, family-based care, home-based care, and afterschool programs). Most systems are now voluntary and linked to financial incentives or receipt of child care subsidies to promote participation (Tout et al., 2010).

Beginning in 2011, the request for applications for the U.S. Department of Education's Race to the Top Early Learning Challenge prioritized the design and implementation of statewide QRISs, leading many states to redouble their efforts to develop one (U.S. Department of Education, 2011). However, policymakers and QRIS administrators struggle with a lack of guidance on which components are the most important to include in order to ensure that ratings represent actual differences in quality (Caronongan, Kirby, Malone, & Boller, 2011). While states differ in how they define and measure child care quality in the QRIS, some domains of quality are common, including licensing compliance (26 states), staff qualifications (26), environment (24), family partnership (24), administration and management (23), and accreditation (21; Caronongan et al., 2011; Tout et al., 2010).[2] Although consensus is growing among experts regarding which components of program quality are most closely related to child development (for example, components of process quality, including supportive teacher–child interactions and use of an evidence-based curriculum; Sabol, Soliday Hong, Pianta, & Burchinal, 2013), agreement is still lacking on how to measure and rate these components in a state-developed QRIS.

Box 1. Approaches for calculating ratings in quality rating and improvement systems

Three primary approaches are used for calculating ratings in quality rating and improvement systems (QRISs): point approach, building block approach, and hybrid approach.

Point approach. Point values are assigned to a variety of quality standards or domains (Zellman & Perlman, 2008), and child care programs or providers earn points for meeting individual quality indicators. Overall quality ratings are assigned based on minimum required points earned from any standard or domain, such as in Wisconsin's QRIS (Tout et al., 2010). The point approach allows states to define priorities for quality areas and gives providers flexibility to focus on certain areas based on their program goals, areas in need of improvement, and financial resources.

Building block approach. The building block approach uses sequential levels that can be reached only by meeting all the criteria within a level and all the levels preceding it (Tout et al., 2010). For example, for a program to move from level 1 to level 2, it must meet all the level 1 standards as well as the level 2 standards (U.S. Department of Health and Human Services, 2010). Building block approaches tend to create more consistency in quality standards because all programs must meet the same standards before advancing to the next rating level (Zellman & Perlman, 2008).

Hybrid approach. Some states use a hybrid approach to ratings, combining elements of both the building block and point approaches. For example, Washington changed its QRIS to a hybrid approach that uses licensing requirements to rate programs at level 1, leadership and management to rate programs at level 2, and points earned in the remaining quality standards to rate programs at levels 3–5 (Washington State Department of Early Learning, 2013).

Even fairly minor alterations to a state's approach for calculating ratings in a quality rating and improvement ssytem can have considerable implications for the distribution of ratings

Research suggests that different approaches for calculating ratings in a quality rating and improvement system can lead to different ratings for some programs, thus affecting the distribution of ratings across the state

Recent studies using nationally representative datasets to simulate QRIS ratings found that the rating calculation approach can considerably alter the distribution of quality ratings (see box 1 for a summary of rating approaches). Specifically, Tout et al. (2014) simulated QRIS ratings using different rating approaches and found that under a building block approach most programs were rated as low quality; under a point approach, many more programs were rated as high quality; and under a hybrid approach, many more programs were rated as moderate to high quality. These results suggest that even fairly minor alterations to a state's calculation approach can have considerable implications for the distribution of ratings.

Michigan's quality rating and improvement system rolled out statewide in 2012, and in 2013 the state changed the approach to calculating ratings

Michigan developed its QRIS, Great Start to Quality, in the early 2000s and rolled out the system statewide in 2012, with more than 2,000 programs participating by the end of the year[3] (see appendix A for more information on Michigan's QRIS). It uses a hybrid rating calculation approach that draws on a self-assessment and independent observations of quality. The ratings range from level 1 to level 5, where level 1 indicates the lowest quality and level 5 indicates the highest quality (see figure A1 in appendix A). Box 2 summarizes

Box 2. Key components and features of Michigan's quality rating and improvement system

- Includes a variety of types of licensed early childhood education programs that serve children from birth to age 5.
- Combines a self-assessment of structural quality and independent observations of process quality, both of which are based on the Program Quality Assessment.
- Uses a hybrid approach that combines aspects of both the point and building block approaches for calculating ratings.
- Encourages statewide participation but is voluntary.
- Publishes ratings so that parents can easily find the best early learning and care options for their children, while educators and providers have clear quality standards toward which to strive.
- Includes incentives for programs to participate (such as staff scholarship tuition assistance, public ratings for high-quality programs, eligibility for some special grant funding opportunities).

In January 2013 Michigan decided to revise its approach to calculating ratings to one that applies different cutscores on the self-assessment and changes the requirements for earning points on the staff qualifications subdomain

the key components of Great Start to Quality, while appendix A provides more detailed information. Table B1 in appendix B summarizes QRISs implemented in the REL Midwest Region.

In January 2013 Michigan decided to revise its approach to calculating ratings as implementation of Great Start to Quality expanded. The changes went into effect in June 2013. Under the initial approach (version 1.0), the requirements for each rating were based on the total number of points on the self-assessment, the number of points in each domain on the self-assessment, and the independent observation of quality. The revised approach (version 2.0) continues to use both the self-assessment and the independent observation of quality but applies different cutscores on the self-assessment and changes the requirements for earning points on the staff qualifications subdomain (see table A1 in appendix A). Version 2.0 still requires a minimum number of points for each domain of the self-assessment but reduces the number of domains in which programs must meet those minimum required points for ratings of levels 2–4 (table 1).

Table 1. Self-assessment rating point requirements in the version 1.0 and 2.0 calculation approaches of Michigan's Great Start to Quality

Rating	Version 1.0	Version 2.0
Level 1	Program must be licensed.	Program must be licensed.
Level 2	Programs must meet minimum required points for a level 2 rating in all five domains.	Programs must meet minimum required points for a level 2 rating in any two domains.
Level 3	Programs must meet minimum required points for a level 3 rating in all five domains.	Programs must meet minimum required points for a level 3 rating in any three domains.
Level 4	Programs must meet minimum required points for a level 4 rating in all five domains.	Programs must meet minimum required points for a level 4 rating in any four domains.
Level 5	Programs must meet minimum required points in all five domains.	Programs must meet minimum required points in all five domains.

Source: Great Start to Quality documentation provided by the Michigan Department of Education's Office of Great Start.

To date, no study of Michigan's rating calculation approach has been conducted, leaving many questions about both approaches under Great Start to Quality. This study provides stakeholders in Michigan with an initial understanding of how the version 1.0 approach classified early childhood education programs according to quality level and how those classifications change under the version 2.0 and other rating approaches. The study is also useful to QRIS administrators and policymakers nationally because it addresses a gap in the literature by comparing how changes to a statewide QRIS affect ratings.

Study approach

This study addresses four research questions:

1. How many programs are rated as low, moderate, and high quality under the version 1.0 approach of Great Start to Quality?

2. How consistent are the self-assessment ratings and ratings on the independent observation of quality under the version 1.0 approach of Great Start to Quality?

3. How are domain scores related to overall self-assessment ratings under the version 1.0 approach of Great Start to Quality?

4. How do the distributions of self-assessment ratings and QRIS scores under Great Start to Quality change with alternative approaches for calculating ratings?

This study is useful to quality rating and improvement system administrators and policymakers nationally because it addresses a gap in the literature by comparing how changes to a statewide quality rating and improvement system affect ratings

Research questions 1–3 provide preliminary data on the landscape of early childhood program quality, based on ratings under the version 1.0 approach of Great Start to Quality. Research question 4 informs key stakeholders in Michigan and other states about how small changes to a rating calculation approach (such as the version 2.0 approach under Great Start to Quality) can influence ratings.

This study draws on multiple data sources collected by state agencies in Michigan (table 2). Appendix C provides additional information on the Self-Assessment Survey and form A of the Program Quality Assessment instruments.

Table 2. Variables from Michigan's Early Childhood Investment Corporation and Michigan's Great Start Collaborative database used in the study

Variable	Type	Range	Source	Research questions addressed
Self-assessment total score	Continuous	0–50	Self-Assessment Survey	1, 2, 3, 4
Self-assessment rating	Ordinal	1–5	Self-Assessment Survey with applied QRIS cutscores	1, 2, 3, 4
Independent observation of quality	Continuous	1.0–5.0	Form A of the Program Quality Assessment	3, 4
QRIS score	Ordinal	1–5	Self-Assessment Survey and form A of the Program Quality Assessment	1, 4

QRIS is quality rating and improvement system.

Source: Materials shared by the Michigan Department of Education's Office of Great Start.

Table 3. Analysis methods and samples for each research question

Research question	Analysis method	Sample of programs
1	Description of the distribution of program ratings.	1,413 programs with a QRIS score (had a self-assessment and completed the independent observation of quality if required); 2,390 programs with only a self-assessment rating.
2	Correlations and comparison of ratings resulting from self-assessment data and independent observation of quality data.	72 programs that had a self-assessment rating and completed the independent observation of quality as of January 2013.
3	Regression analysis to examine the strength of the relationships between each domain score in the self-assessment and the overall self-assessment rating.	2,390 programs with a self-assessment rating.
4	Comparison of self-assessment ratings and QRIS scores, resulting from different rating calculation methods and alternative cutscores.	2,390 programs with a self-assessment rating; 72 programs that had a self-assessment rating and completed the independent observation of quality.

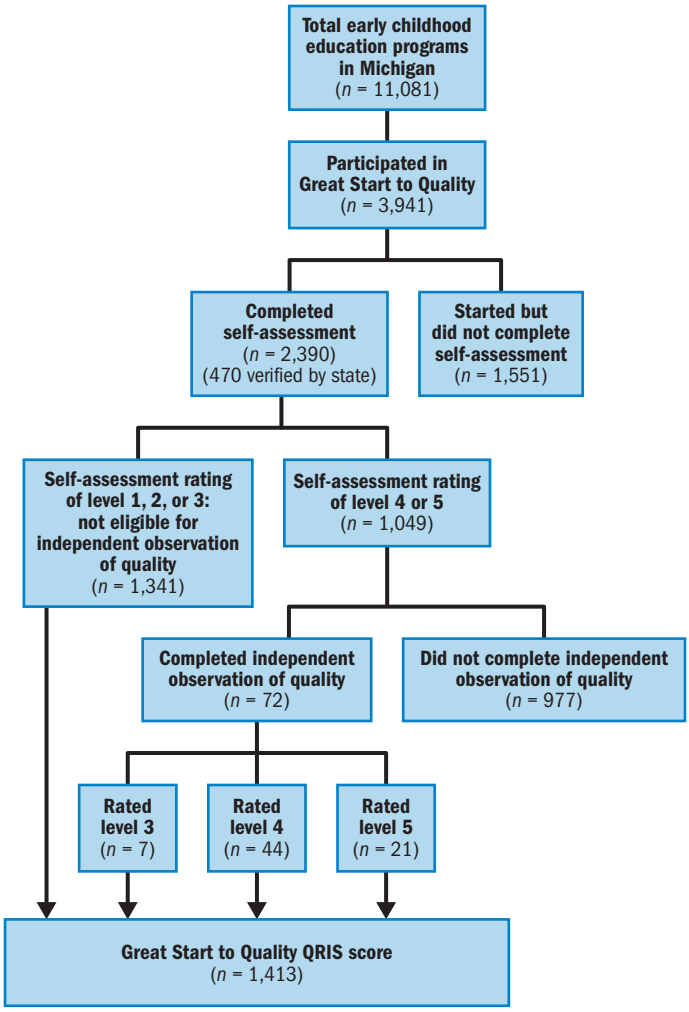
Source: Authors.

This study employed different quantitative analysis methods to address each research question using a sample of programs (ranging from 72 to 2,390) with the specific combination of data needed for the analysis (table 3; see appendix D for detailed information about analysis methods).

Figure 1 depicts the stages of rating participation and completion among the 11,081 licensed early child care programs according to the data provided by Michigan as of January 16, 2013 (see appendix E for more information on the programs that were part of this study). Among these licensed programs, 3,941 participated voluntarily in Great Start to Quality by beginning the self-assessment; 2,390 of those completed it. Among the 2,390 programs that completed the self-assessment, 470 were verified by a state agency representative who reviewed supporting documentation to check the accuracy of the self-assessment ratings. In cases where the verified rating differed from the self-assessment rating, the state used the verified rating (see appendix E for information on how many program ratings changed due to state verification).

For the 1,341 programs with a self-assessment rating of level 1, 2, or 3, this was their QRIS score. The 1,049 programs with a self-assessment rating of level 4 or 5 were eligible for the independent observation of quality to achieve a QRIS score of level 4 or 5. The independent observation of quality requires that a state-approved rater conduct classroom observations in-person for the program; only 72 programs (6.9 percent of those eligible) completed the independent observation of quality by January 2013 (see appendix E for more information on these programs). The score on the independent observation of quality overrides the self-assessment rating in determining the QRIS score, though programs must have high ratings for both to receive a QRIS score of level 4 or 5. Programs that self-rate as high but have a low score on the independent observation of quality (below 3.5) receive a QRIS score of level 3. The total number of programs with a QRIS score was 1,413 (738 licensed center-based programs and 675 family child care homes).

Figure 1. Number of programs in Michigan at each stage of participation in the quality rating and improvement system



QRIS is quality rating and improvement system.

Note: "Self-assessment" is the Self-Assessment Survey instrument used for self-ratings. "Independent observation of quality" is form A of the Program Quality Assessment instrument. Data are as of January 16, 2013.

Source: Data provided by the Michigan Department of Education's Office of Great Start.

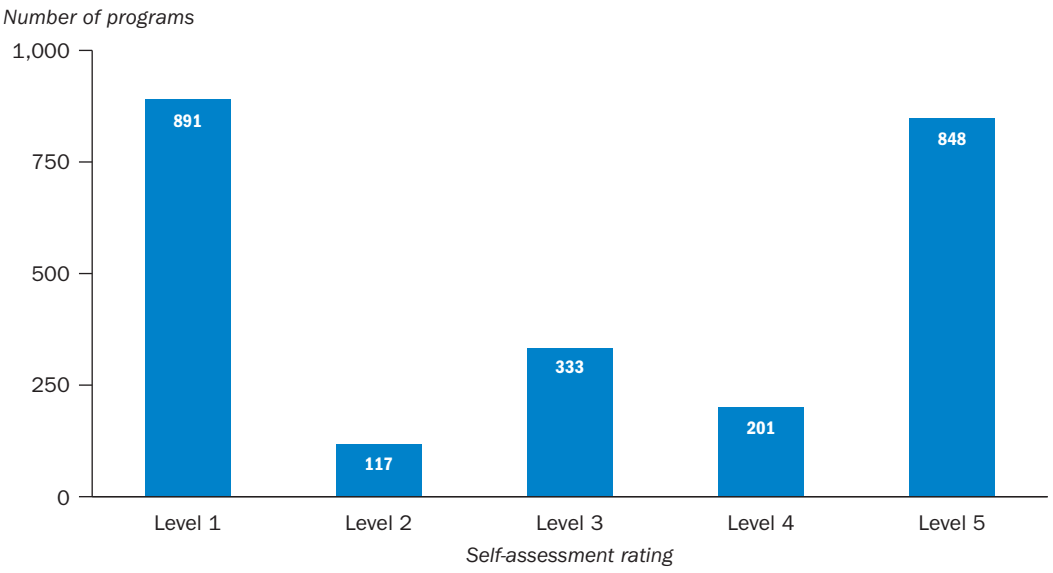
What the study found

This section details the main findings of the study.

Almost three-quarters of programs self-rated at either the lowest or highest levels of quality

Under Great Start to Quality's version 1.0 approach to calculating ratings, programs tended to self-rate at the lowest (1) and highest (5) levels (figure 1). Of the 2,390 programs that completed the self-assessment, 891 (37.2 percent) self-rated at level 1, 848 (35.5 percent) self-rated at level 5, and 651 (27.2 percent) self-rated at level 2, 3, or 4 (figure 2). This distribution is consistent with research on other QRISs, suggesting that only a small percentage of programs have ratings in the middle range of quality (Caronongan et al., 2011).

Figure 2. Most programs self-rated at the lowest and highest levels of quality under the version 1.0 approach of Michigan's Great Start to Quality



n = 2,390.

Note: Data are as of January 16, 2013.

Source: Data provided by the Michigan Department of Education's Office of Great Start.

Quality rating and improvement system scores that incorporated both self-assessment ratings and an independent observation of quality resulted in many fewer programs rated as high quality

Self-assessment ratings and QRIS scores are not necessarily the same in Great Start to Quality. The QRIS score incorporates both the self-assessment rating and the independent observation of quality. Of the 1,413 programs with a QRIS score (those with a self-assessment rating of level 1, 2, or 3 and those with a self-assessment rating of level 4 or 5 that also received an independent observation of quality), 891 (63.1 percent) had a QRIS score of level 1, and 65 (4.6 percent) had a QRIS score of level 4 or 5 (figure 3).

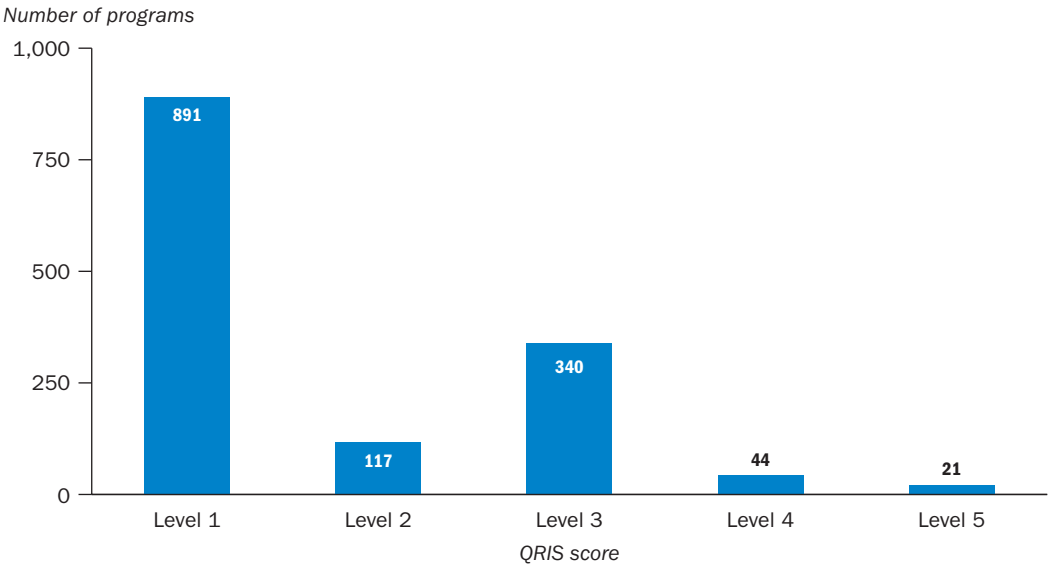
Under Great Start to Quality's version 1.0 approach to calculating ratings, programs tended to self-rate at the lowest (1) and highest (5) levels

Of the 72 programs that received an independent observation of quality (those with a self-assessment rating of level 4 or 5), 65 received a QRIS score of level 4 or 5, while 7 received a QRIS score of level 3 (table 4). The majority of programs that self-rated at level 4 or 5 (977) were on a wait list for an independent observation as of January 2013. The small percentage of programs with a high QRIS score is due mainly to the low number of programs that received an independent observation of quality and does not necessarily indicate a lack of high-quality programs (see appendix F for the descriptive statistics of programs at each level). Thus, the distribution of QRIS scores should be considered preliminary and interpreted with caution. More data are needed to understand whether patterns will shift when more programs have an independent observation of quality.

Ratings on the self-assessment and independent observations of quality differed for more than half of programs with both types of ratings

Of the 72 programs with both a self-assessment rating and an independent observation of quality under the version 1.0 approach of Great Start to Quality, 37.5 percent had the

Figure 3. Most programs had a QRIS score of level 1 under the version 1.0 approach of Michigan's Great Start to Quality, and very few had a QRIS score of level 4 or 5



QRIS is quality rating and improvement system.

n = 1,413.

Note: Data are as of January 16, 2013.

Source: Data provided by the Michigan Department of Education's Office of Great Start.

Table 4. Roughly 63 percent of self-assessment ratings were higher than ratings on the independent observation of quality under the version 1.0 approach of Michigan's Great Start to Quality

Self assessment rating	Rating on the independent observation of quality		
	Level 3 (*n* = 7)	Level 4 (*n* = 44)	Level 5 (*n* = 21)
Level 4 (*n* = 9)	3 (4.2 percent)	**6 (8.3 percent)**	0 (0.0 percent)
Level 5 (*n* = 63)	4 (5.5 percent)	38 (52.7 percent)	**21 (29.2 percent)**

n = 72.

Note: Bolded cells indicate that self-assessment ratings and ratings on the independent observations of quality were the same. Data are as of January 16, 2013.

Source: Authors' calculations based on data provided by the Michigan Department of Education's Office of Great Start.

Of the 1,413 programs with a quality rating and improvement system score, very few had a score of 4 or 5

same rating on both, and the rest had a rating on the independent observation that was one or more levels lower than the self-assessment rating (see table 4). More than half of programs had a rating on the independent observation that was one level lower than the self-assessment rating: 52.7 percent had a self-assessment rating of level 5 and a level 4 rating on the independent observation, and 4.2 percent had a level 4 self-assessment rating and a level 3 rating on the independent observation. Four programs (5.5 percent) that had a level 5 self-assessment rating had a level 3 rating on the independent observation. No programs had a rating on the independent observation that was higher than the self-assessment rating. There was also no significant association between programs' total scores

on the Self-Assessment Survey and form A of the Program Quality Assessment (Spearman's $\rho = .188$, $p = .114$), which may suggest that the definition of quality in the two instruments is not the same (see appendix G for more detail on the nonparametric measures of association). However, as stated above, the findings that incorporate data on the independent observation should be considered preliminary and interpreted with caution because so few programs have relevant data.

No single domain drove the overall self-assessment rating

To test whether any of the five domains on the Self-Assessment Survey drove the overall self-assessment ratings, generalized ordered logistic regression was used to test the relationship between each domain score and programs' self-assessment ratings. The results are reported as odds ratios for each rating level increase (from level 1 to level 2, from level 2 to level 3, and from level 3 to level 4 or level 5) for each of the five domains. Odds ratios represent the likelihood that a program would have a higher overall self-assessment rating if the domain score increased by one point. A statistically significant odds ratio above 1 indicates an increased likelihood of a higher overall self-assessment rating, and a statistically significant odds ratio below 1 indicates a decreased likelihood. For example, the odds ratio of 1.62 for staff qualifications indicates that an additional point on the staff qualifications domain score increases by 62.0 percent the odds of moving from level 1 to level 2. Both an odds ratio of 1 and an odds ratio that is not statistically significant indicate no change in likelihood.

Self-assessment ratings and independent observations of quality were not significantly correlated, suggesting that they measure different aspects of quality

The odds ratios are significant and positive for all domains at all rating levels, except for the family and community partnerships domain, for which there was no relationship with moving from level 1 to level 2 or with moving from level 2 to level 3 (table 5). These analyses thus suggest that no one domain drives the overall self-assessment rating.

In general, additional domain points tend to have a smaller impact on the chance of moving up at the lower rating levels (moving from level 1 to level 2 and from level 2 to level 3), compared with moving from level 3 to level 4 or level 5. This can be seen in the

Table 5. All domains contributed to the overall self-assessment rating under the version 1.0 approach of Michigan's Great Start to Quality, especially at the highest levels

	Odds ratio		
Self assessment rating domain	Moving from level 1 to level 2	Moving from level 2 to level 3	Moving from level 3 to level 4 or level 5
Administration and management	1.35***	1.39***	2.32***
Curriculum	1.14*	2.05***	1.78***
Environment	1.23***	1.38***	1.93***
Family and community partnerships	1.01	1.08	2.37***
Staff qualifications and professional development	1.62***	1.74***	2.22***

* is significant at $p < .05$; *** is significant at $p < .001$.

Note: Data are as of January 16, 2013.

Source: Authors' calculations based on data provided by the Michigan Department of Education's Office of Great Start.

larger odds ratios for moving from level 3 to level 4 or level 5 than for moving from level 1 to level 2 or from level 2 to level 3 (see appendix D for information on these analyses).

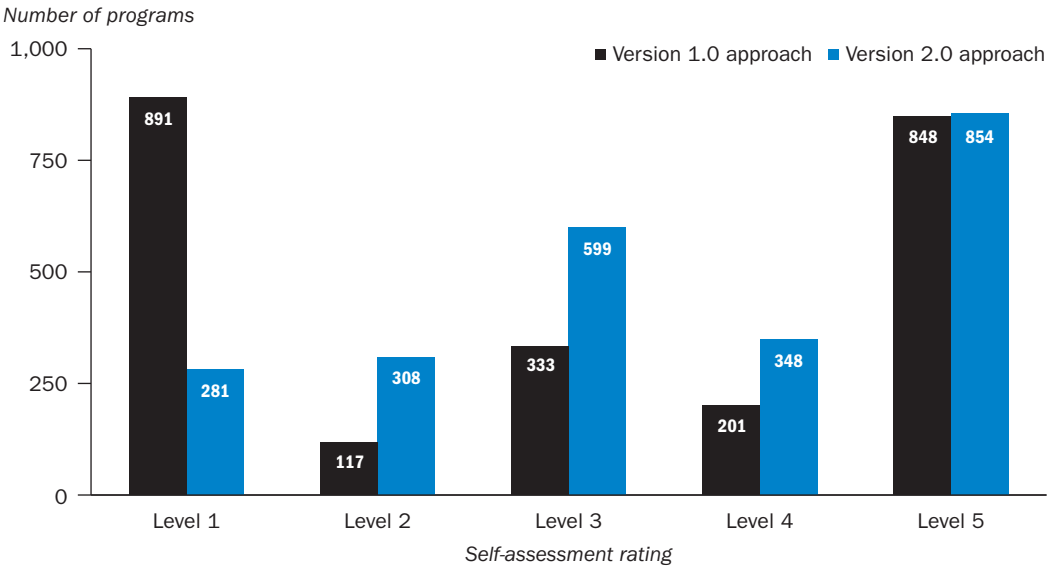
Programs had higher ratings under the version 2.0 and total score approaches than under the version 1.0 approach of Great Start to Quality

This section compares program ratings under the version 1.0 approach with those based on three alternative scoring approaches: the version 2.0 approach, a total score approach, and an approach that changed the cutscores for the independent observation of quality. Because only 72 programs had an independent observation of quality as of January 2013, these analyses focus primarily on the self-assessment ratings.

The version 2.0 approach makes it easier for programs to be rated as moderate quality, reduces the number of programs rated as low quality (level 1), and maintains the same number of programs rated as high quality (level 5)

Programs had a higher self-assessment rating under the version 2.0 approach than under the version 1.0 approach, with some programs increasing by two or more levels. Of the 2,390 programs that completed the self-assessment, 891 (37.3 percent) self-rated at level 1 under the version 1.0 approach, compared with 281 (11.8 percent) under the version 2.0 approach (figure 4 and table 6). The number of programs at levels 2, 3, and 4 each increased by more than 100 from the version 1.0 approach to the version 2.0 approach. The number of programs that qualified for a level 5 rating did not change substantially (from 848, or 35.4 percent, to 854, or 35.7 percent) because the criteria for this highest rating level was essentially the same under both approaches. The median rating was level 3 under the version 1.0 approach and level 4 under the version 2.0 approach, suggesting that the version 2.0 approach makes it easier for programs to be rated as moderate quality,

Figure 4. Self-assessment ratings under the version 2.0 approach of Michigan's Great Start to Quality included fewer programs at level 1 and more at levels 2, 3, and 4 than under the version 1.0 approach



n = 2,390.

Note: Data are as of January 16, 2013.

Source: Authors' calculations based on data provided by the Michigan Department of Education's Office of Great Start.

Table 6. Change in self-assessment ratings under the version 1.0 and 2.0 approaches of Michigan's Great Start to Quality

Rating under version 1.0 approach	Rating under version 2.0 approach				
	Level 1 (*n* = 281)	Level 2 (*n* = 308)	Level 3 (*n* = 599)	Level 4 (*n* = 348)	Level 5 (*n* = 854)
Level 1 (*n* = 891)	**281**	291	279	34	6
Level 2 (*n* = 117)	0	**17**	87	13	0
Level 3 (*n* = 333)	0	0	**233**	100	0
Level 4 (*n* = 201)	0	0	0	**201**	0
Level 5 (*n* = 848)	0	0	0	0	**848**

Note: Bolding indicates ratings that did not change between versions 1.0 and 2.0. Data are as of January 16, 2013.

Source: Authors' calculations based on data provided by the Michigan Department of Education's Office of Great Start.

reduces the number of programs rated as low quality (level 1), and maintains the same number of programs rated as high quality (level 5).

As many as 235 of the 810 programs that moved up from level 1, including all 6 that moved from level 1 to level 5, did so at least in part by no longer having to earn points in three separate subdomains of the staff qualifications domain, a requirement under the version 1.0 approach that was not included under the version 2.0 approach (see table A1 in appendix A). All other increases can be attributed to no longer having to meet the minimum points in every Self-Assessment Survey domain.

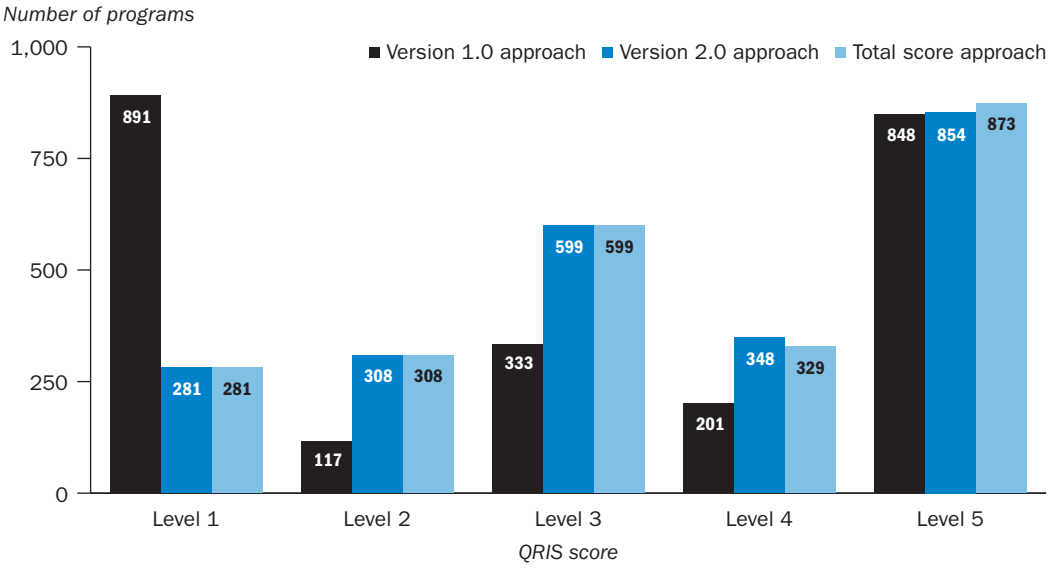
If the state relied solely on self-assessment total scores and removed the domain score requirements, ratings would be nearly identical to those under the version 2.0 approach

The distribution of self-assessment ratings under the total score approach was almost identical to that under the version 2.0 approach. Another calculation approach is to rely only on self-assessment total scores, with no domain score requirements. Michigan expressed an interest in understanding the implications of a rating approach that does not include domain score requirements. If the state relied solely on self-assessment total scores and removed the domain score requirements, more programs would have a higher self-assessment rating than under version 1.0 and version 2.0 (figure 5). The only difference between the distribution of ratings under the version 2.0 approach and the total score approach is that 19 programs would move from level 4 under the version 2.0 approach to level 5 under the total score approach.

Changing the cutscores on the independent observation of quality by as little as 0.10 point would alter the distribution of QRIS scores. Another way to revise the calculation approach is to change the cutscores on the independent observation of quality, which range from 0 to 5 and increase in increments of 0.01. Under the version 1.0 and 2.0 approaches, cutscores are below 3.5 for level 3, 3.5–4.49 for level 4, and 4.5 and higher for level 5.

Dropping cutscores by as little as 0.10 point would increase the percentage of programs rated at level 5 from 29.2 percent to 34.8 percent (table 7). Increasing cutscores by 0.10 point would decrease the percentage of programs rated at level 5 to 25.0 percent. However, these findings are based on the small sample of programs (72) with an independent observation of quality, and further analyses should be conducted when the state has completed more independent observations to determine whether the pattern holds.

Figure 5. Self-assessment ratings under the total score approach were higher than under the version 1.0 approach of Michigan's Great Start to Quality and similar to those under the version 2.0 approach



n = 2,390.

Note: Data are as of January 16, 2013.

Source: Authors' calculations based on data provided by the Michigan Department of Education's Office of Great Start.

Table 7. Changing the cutscores on the independent observation of quality alters the distribution of QRIS scores from that under the version 1.0 and 2.0 approaches of Michigan's Great Start to Quality

Change in independent observation of quality cutscore	Number of programs at each QRIS score		
	Level 3	Level 4	Level 5
0.25 point lower	3 (4.1 percent)	37 (51.5 percent)	32 (44.4 percent)
0.10 point lower	3 (4.1 percent)	44 (61.1 percent)	25 (34.8 percent)
No change	7 (9.7 percent)	44 (61.1 percent)	21 (29.2 percent)
0.10 point greater	8 (11.1 percent)	46 (63.9 percent)	18 (25.0 percent)
0.25 point greater	13 (18.1 percent)	50 (69.4 percent)	9 (12.5 percent)

QRIS is quality rating and improvement system.

n = 72.

Note: Data are as of January 16, 2013.

Source: Authors' calculations based on data provided by the Michigan Department of Education's Office of Great Start.

Discussion

Despite making strides in rolling out a QRIS statewide, Michigan is still in the early stages of the process. In the first year and a half of implementation of the version 1.0 approach, 43.8 percent of programs had self-assessment ratings at the highest levels (4 and 5). However, when focusing on just the QRIS scores (based on the self-assessment rating and the rating on the independent observation of quality), fewer than 5.0 percent of participating programs had a QRIS score of level 4 or 5, largely because of challenges in completing the independent observations. Among programs eligible for the independent observation, only 6.9 percent had completed it by January 2013, partly because state resources were limited (for example, limited funds to train observers and administer the independent observation) and partly because not all programs chose to participate.

More independent observations of quality are needed to understand how many programs will ultimately have the highest quality rating and improvement system score

In spring and summer 2013 the state increased efforts to administer the independent observation for eligible programs. This included hiring staff to administer the observations and conducting ongoing training to ensure that observers rate program quality reliably. After the independent observations have been completed for most (if not all) eligible programs, future studies should examine the overall distribution of QRIS scores with the new data.

Self-assessments and independent observations of quality did not always result in the same rating for the limited number of programs with both types of data

Self-assessment ratings tended to be higher than those on the independent observation of quality for the 72 programs with both types of ratings. Ratings differed for more than 60 percent of programs with both types of ratings, and for all programs with different ratings, the self-assessment rating was higher than the rating on the independent observation of quality, though most differed by only one level. More independent observations of quality must be conducted for eligible and interested programs before the state can determine whether self-assessment ratings and independent observations are assessing quality in the same or a different way.

As Great Start to Quality was developed, the state's QRIS advisory team chose to define early childhood education program quality based on both structural and process quality. Programs identified as high quality based on structural quality alone (those with only self-assessment ratings) may look very different from programs identified as high quality based on a combination of structural and process quality (those with ratings on both the self-assessment and the independent observation of quality). Although structural quality tends to predict process quality (Burchinal, Cryer, Clifford, & Howes, 2002; National Institute of Child Health and Human Development Early Child Care Research Network, 2002; Pianta et al., 2005), the relationship is not perfect, and structural quality ratings should not be expected to align precisely with process quality ratings.

Great Start to Quality was intentionally designed to incorporate measures of process quality for programs with high ratings based on structural quality in order to ensure that the highest QRIS scores were restricted to programs that have a strong structural foundation and provide high-quality interactions and instruction. The self-assessment ratings

Programs identified as high quality based on structural quality alone may look very different from programs identified as high quality based on a combination of structural and process quality

and the ratings from the independent observation of quality are likely to differ somewhat, but that might reflect the added value of assessing process quality rather than a problem of poor alignment among the instruments.

The differences also could be due partly to the different modes of data collection. The self-assessment ratings are reported by program directors or other program staff (and verified by the state for some programs), while the independent observation of quality is conducted by outside observers hired by the state. Self-reported data could be vulnerable to bias, intentional or not, inflating the results because the staff completing the self-assessment has a vested interest in receiving a higher QRIS score. More than 40 percent of programs with both self-assessment ratings and state-verified ratings had different ratings (see figure E1 and table E5 in appendix E), but for almost a third the self-assessment rating was lower than the state-verified rating. This suggests that self-reported data are vulnerable to bias, but the direction is not always in favor of the program.

Further research is needed to test the relationship between self-assessment ratings and ratings on the independent observations and to study the extent to which differences in the ratings are due to differences in the constructs measured by the instruments, differences in the mode of data collection, or other factors. Again, these findings should be interpreted with caution due to the limited sample of programs with both a self-assessment rating and an independent observation of quality.

All five self-assessment domains contribute to the overall self-assessment rating under the version 1.0 approach of Great Start to Quality, but whether that will remain true under the version 2.0 approach is unknown

Michigan officials expressed interest in understanding whether any one domain was driving overall self-assessment ratings under Great Start to Quality. The logistic regression analyses under the version 1.0 approach found no evidence of this; rather, scores on all domains were predictive of increases in self-assessment ratings, especially for highly rated programs. These findings do not have implications for individual programs' efforts in quality improvement, because the specific improvements needed to reach the next rating level are always determined by programs' own strengths and weaknesses on each self-rated domain.

Although each domain is associated with the overall self-assessment rating under the version 1.0 approach, how these relationships will change under the version 2.0 approach is unclear. The point and threshold structure of the version 1.0 approach meant that providers needed to meet a minimum number of points in every domain to reach the next rating level. However, under the version 2.0 approach providers will no longer have to meet point requirements in every domain to receive a rating of level 2, 3, or 4. Thus it remains to be seen whether programs will make improvements only in domains in which it is easier for them to earn additional points. That would cause the relationship between each domain and the overall self-assessment rating or QRIS score to be very different.

The shift away from requiring a minimum number of points in all quality domains under the version 2.0 approach implies that there will be greater diversity in the type of quality exhibited by the programs at each rating level. The state's primary motivation in changing the approach was to remove barriers that prevented programs from achieving a rating

Under the version 2.0 approach providers will no longer have to meet point requirements in every domain to receive a rating of level 2, 3, or 4; the shift away from requiring a minimum number of points in all quality domains under the version 2.0 approach implies that there will be greater diversity in the type of quality exhibited by the programs at each rating level

above level 1 rather than to redefine quality measured. However, ratings of levels 2, 3, and 4 will no longer necessarily align with the multidimensional definition of quality that the state drew on in its initial design. Furthermore, the changes will also shift the incentives for programs to make quality improvements as they work toward higher ratings. Michigan should continue to study the relationship between domain scores and self-assessment ratings after the version 2.0 approach has been fully implemented to determine how much the change affects the influence of each domain. Other states that are considering changing their QRIS to make higher ratings more achievable might also consider how these changes will affect the type of quality measured.

Reduced domain score requirements mean that programs tended to have higher self-assessment ratings under the version 2.0 approach to Great Start to Quality

The distribution of self-assessment ratings under the version 2.0 approach of Great Start to Quality, which reduces domain score requirements, and the total score approach, which eliminates domain score requirements completely, was almost identical

The results of these analyses suggest that changes in the rating calculation approach can influence the number of programs at each rating level. Michigan began implementing the version 2.0 approach in June 2013, and a large percentage of programs would have a higher self-assessment rating under the version 2.0 approach. These findings are consistent with other research suggesting that the distribution of ratings depends heavily on the calculation approach (Tout et al., 2014).

As states move forward with implementing their QRISs, careful consideration should be given to the calculation approach and the ways that even small differences in the criteria for determining ratings can affect both the total distribution of ratings and the ratings of individual programs. More research is needed to determine whether the version 2.0 approach leads to wider differences between self-assessment ratings and ratings on the independent observation of quality and to examine how the version 2.0 approach compares with the point, building block, and hybrid approaches used by other states. Research also is needed to understand the validity of the version 2.0 approach in terms of relationships with other quality measures and the prediction of child outcomes.

Shifting to the version 2.0 approach of Great Start to Quality yields essentially the same distribution of self-assessment ratings as a total score approach that eliminates domain score requirements completely

The distribution of self-assessment ratings under the version 2.0 approach of Great Start to Quality, which reduces domain score requirements, and the total score approach, which eliminates domain score requirements completely, was almost identical. Although the version 2.0 approach still has some domain score requirements, the requirements were loosened such that the distribution of quality is similar to the total score approach, in which the domain requirements are removed altogether. Given this finding, Michigan could achieve similar self-assessment rating results by either reducing the number of domains that have cutscores or removing all domain score requirements. Using total scores alone would reduce the complexity of the QRIS and make it easier for the state to calculate ratings and for programs to understand the system, while resulting in essentially the same distribution of ratings as under the version 2.0 approach.

However, both the version 2.0 and total score approaches change the definition of quality signaled by the self-assessment ratings, compared with the version 1.0 approach. Under the version 1.0 approach programs were required to meet standards in five domains at each

QRIS level; under the version 2.0 approach these requirements have been loosened considerably, especially at the lower QRIS rating levels. Programs with a rating of level 4 or 5 still have to meet most or all domain score requirements from the self-assessment rating and demonstrate high process quality on the independent observation of quality; however, programs with a rating of level 2 or 3 can choose multiple domains in which they do not have to meet any requirements. Quality signaled by a rating of level 2 or 3 is thus much broader and looser in its definition under the version 2.0 approach or another total score approach than under the version 1.0 approach.

Michigan and other states must therefore consider the tradeoffs of using an approach that reduces or eliminates domain score requirements. Each state must decide which is more important—the benefit of using a simple calculation system or the benefit of defining quality as multidimensional.

Each state must decide which is more important—the benefit of using a simple calculation system or the benefit of defining quality as multidimensional

Shifting to the version 2.0 approach of Great Start to Quality may have cost implications

Michigan's revised QRIS calculation approach may have cost and capacity implications: more programs will self-rate at level 4 (and level 5 in a few cases), increasing the number eligible for the independent observation of quality. Administering the independent observation has been costly for Michigan, particularly because of the labor costs of supporting independent observers, and the state has faced delays in completing observations for all programs eligible under the version 1.0 approach. Increasing the number of programs eligible for the independent observation could exacerbate these challenges.

However, Michigan now uses an alternate pathway to high ratings for some programs, including the state's Great Start to Readiness and Head Start programs, as well as programs with accreditation by the National Association for the Education of Young Children or the National Association for Family Child Care. Programs in these categories account for about a third of programs rated in January 2013 (see appendix E). Because accreditation requirements and Great Start to Readiness and Head Start program standards are comparable to requirements for higher ratings under Great Start to Quality, the state now exempts this subset of programs from the independent observation of quality and uses only self-assessment ratings for a level 4 rating (programs still have to participate in the independent observation to receive a level 5 rating). This reduces the burden of conducting independent observations for eligible programs that rate themselves at level 4. Also, some programs in Michigan already use the independent observation instrument for other purposes in their accountability systems, such as for self-assessment or to meet funding requirements, and the state is examining how to leverage the data collected by programs for use with Great Start to Quality.

Study limitations

This section describes three limitations of the current study.

First, the data available as of January 2013 were limited because the state's QRIS was fairly new. Michigan had completed just 72 independent observations of quality, even though 1,049 programs were eligible to receive one based on their self-assessment ratings. Furthermore, programs that completed the independent observation were not randomly sampled and were disproportionately representative of certain locations in the state (see

appendix D). Because of these limitations, caution is advised in interpreting findings based on analyses using the independent observations.

Second, the self-assessment ratings are based on self-report for 80 percent of programs and state-verified ratings for 20 percent of programs. This suggests some concerns about the reliability of the self-assessment rating data, as different modes of data collection were used for the two types of ratings. Furthermore, for more than 40 percent of programs with both types of ratings, the self-assessment ratings differed from the state-verified ratings. This suggests that the data for programs with only self-reported data may be biased.

Third, because Great Start to Quality is still voluntary—albeit with incentives to participate—the sample of programs and providers included in the analyses is not representative of all early childhood education programs in Michigan, only of those rated in the QRIS. Differences between rated and nonrated programs may exist; however, because of the limited data on programs that are not rated, extensive comparisons of quality were not possible. Examining basic program characteristics such as license type and total enrollment revealed some differences between participating and nonparticipating programs (see appendix E).

Validating a state's quality rating and improvement system can lend the system credibility and enhance trust in the system ratings for programs themselves, policymakers, and families as they make important child care decisions

Future directions

This study's findings suggest that there is still a need to build an understanding of the psychometric properties of the data collection instruments used in QRISs—including validity and reliability. In Michigan this could include a study of the psychometric properties of the Self-Assessment Survey used for the self-assessment ratings and how the Self-Assessment Survey relates to other valid and reliable self-report measures of program quality. Because the self-assessment rating is a home-grown measurement system initially based on form B of the Preschool Program Quality Assessment, it is now ripe for a validation study to determine its utility in assessing quality. Further research into the properties of instruments in QRISs across states and how they function in this high-stakes setting is warranted.

Also, as states invest substantial federal and state funds into developing QRISs to rate and improve the quality of early childhood education, there is a need to validate the systems themselves. States that have been awarded Race to the Top Early Learning Challenge grants have federal funds allocated for developing and validating statewide QRISs, and during the next five years many more validation studies will be conducted across the country. Validating a QRIS is a complex, multistep process whereby states assess the degree to which their established ratings reliably measure the differences in program quality across the state. Validation studies are necessary because ratings are central to the overall QRIS and because it is important to document that the ratings are both accurate and meaningful (Zellman & Fiene, 2012). In addition, validating a state's QRIS can lend the system credibility and enhance trust in the system ratings for programs themselves, policymakers, and families as they make important child care decisions.

Validation is important because common rating components and various composite rating calculation approaches within QRISs are only tenuously connected to student learning outcomes (Sabol et al., 2013). Sabol et al. (2013) simulated QRIS rating approaches in 11 states using data on prekindergarten programs and children from two large studies conducted during 2001–04 that focused on quality but not specifically on QRISs. The

researchers found that process quality characteristics (such as adult–child interactions and learning environment) were linked most closely with child outcomes and that structural characteristics (such as staff qualifications, class size, and family–community partnerships) had weaker or inconsistent relationships with child outcomes. They also found that QRISs that combine both process and structural components into one overall rating had only weak relationships with student outcomes. Thus, one recommendation to states as they consider changes to their QRISs is to focus more closely on the quality domains that are closest to student success.

Finally, many in the field are calling for simpler QRISs with fewer rating options, spurred in part by Sabol et al. (2013). This is something for all states to consider as they revise their calculation systems. Simpler systems may be more useful for families, policymakers, educators, and community members and more likely to accurately differentiate program quality and child outcomes. However, recent research suggests that the components of QRISs that are most closely linked to student outcomes are the expensive, observational measures of process quality (Sabol et al., 2013). Therefore, simpler rating systems may not necessarily be the less expensive options for states.

Appendix A. Description of Michigan's quality rating and improvement system, Great Start to Quality

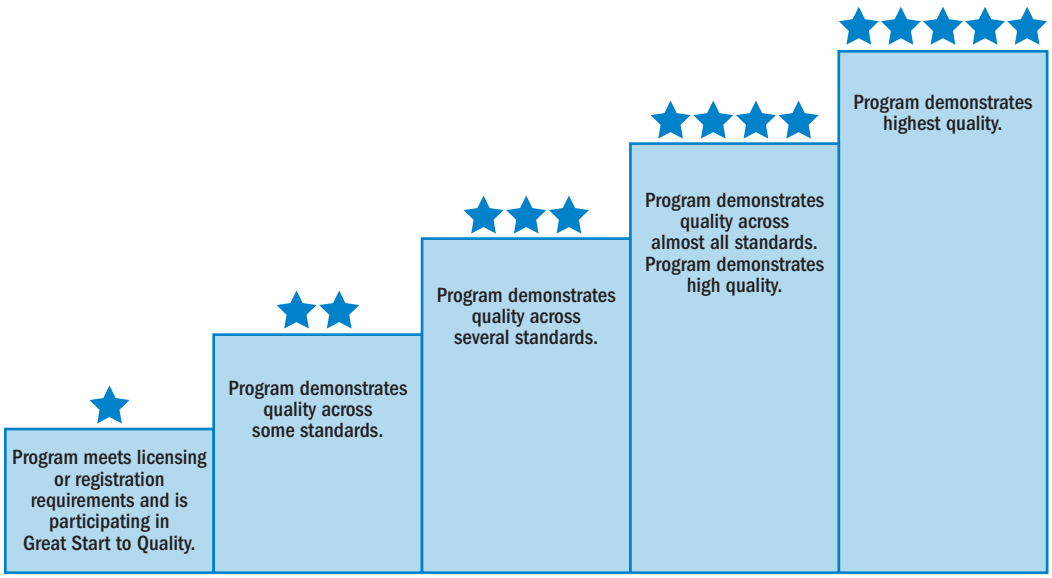
Michigan's development of a quality rating and improvement system (QRIS) began in 2005 when a small group of early childhood leaders developed a set of recommendations for a quality rating system for licensed child care providers. In 2008 the Early Childhood Investment Corporation became involved and formed an advisory team to revisit the design and implementation of Michigan's QRIS. Piloting began in 2009 in four Detroit neighborhoods. In 2010 the Early Learning Advisory Council, in collaboration with the Early Childhood Investment Corporation and the BUILD Initiative, convened stakeholders to re-examine the 2007 published recommendations for a QRIS. Michigan invested $8 million in American Reinvestment and Recovery Act of 2009 funds to create the state's current QRIS for licensed early childhood education programs, including child care centers, licensed family child care homes, Head Start programs, state preschool programs, and other types of programs serving children from birth to age 5.[4] In 2012 the new QRIS, Great Start to Quality, was implemented statewide on a voluntary basis.

Great Start to Quality uses a hybrid rating calculation approach that draws on self-assessment ratings and an independent observation of quality. The ratings range from 1 (lowest quality) to 5 (highest; figure A1).

Great Start to Quality uses both self-reported and independently observed quality instruments, covering different aspects of program quality

As Great Start to Quality was developed, the state's QRIS advisory team chose to define early childhood education program quality according to the domains measured in High/Scope's Program Quality Assessment instrument. The instrument has two components: form B measures structural quality and form A measures process quality.

Figure A1. Descriptions of program quality using the five-level rating system of Michigan's quality rating and improvement system



Source: Adapted from QRIS National Learning Network (2007).

Self-Assessment Survey (structural quality). The Self-Assessment Survey measures five domains of early childhood education quality: administration and management, curriculum, environment, family and community partnerships, and staff qualifications and professional development. Scores range from 0 to 50 points, with points awarded in increments of 1 or 2 for each quality standard achieved by a program in each domain. Item scores are summed to calculate domain scores, and domain scores are summed to calculate the overall self-assessment rating (see tables C1, C4, and C5 in appendix C for more information on domains, subdomains, and scoring). All programs participating in Great Start to Quality complete the Self-Assessment Survey online and upload supporting documentation. It is designed to be a self-reported measure, although the state verified some ratings during the first year of implementation (see figure E1 and table E5 in appendix E for a comparison of the self-reported and state-verified ratings).[5] Validity and reliability data are not available for the Self-Assessment Survey as implemented in Michigan. However, the Self-Assessment Survey was developed to include the constructs in form B of the Program Quality Assessment, a valid and reliable measure of program quality that focuses primarily on structural quality. (See table C3 in appendix C for more information on items included in form B of the Program Quality Assessment.)[6]

Preschool Program Quality Assessment (process quality). For the independent observation of quality, Michigan uses form A of the Program Quality Assessment. Form A measures four domains: adult–child interaction, curriculum planning and assessment, daily routine, and learning environment. It has multiple models for home-based and center-based programs serving infants, toddlers, and preschoolers. Items are scored on a 1–5 scale based on which criteria programs meet on the item, and the item scores are averaged to calculate the overall independent observation score, which ranges from 1.0 to 5.0. Under Great Start to Quality, only programs with a self-assessment rating of level 4 or 5 can voluntarily participate in the independent observation of quality, so independent observations are available only for programs that have high structural quality according to the self-assessment rating. If a program has more than one classroom, all classrooms in the program are observed and their scores averaged. For the 72 programs that completed the independent observation, more than 300 classroom observations were conducted. (See table C2 in appendix C for more information on items included in form A of the Program Quality Assessment, table C3 for items on form B, table C4 for a comparison of ratings on the self-assessment and independent observation of quality, and table C5 for more information on domains, subdomains, and scoring.)

Michigan's Great Start to Quality uses a hybrid rating calculation approach based on self-assessment ratings and independent observations of quality

In developing Great Start to Quality, Michigan intentionally combined scores from two instruments, the Self-Assessment Survey and form A of the Program Quality Assessment, so that programs could have a high QRIS score only if they had high ratings in both structural and process quality. Great Start to Quality is a hybrid because it combines elements of both a point approach and a building block approach. It uses a multistep process to determine the QRIS score for each participating licensed program. To begin the process, licensed programs complete the Self-Assessment Survey and are assigned an overall self-assessment rating of level 1–5 based on their preliminary score and minimum point requirements for domain scores (table A1; see appendix E for information on state verification of scores for some programs).

Table A1. Scoring requirements on the Self-Assessment Survey of Michigan's Great Start to Quality, by domain and scoring approach

	Version 1.0 approach					Version 2.0 approach					Total score approach				
		QRIS score					QRIS score					QRIS score			
Domain	Total points	Level 2	Level 3	Level 4	Level 5	Total points	Level 2	Level 3	Level 4	Level 5	Total points	Level 2	Level 3	Level 4	Level 5
Administration and management	6	2	4	4	4	6	2	4	4	4	6	na	na	na	na
Curriculum	12	4	6	8	8	12	4	6	8	8	12	na	na	na	na
Environment	8	2	4	6	6	8	2	4	6	6	8	na	na	na	na
Family and community partnerships	8	4	4	6	6	8	4	4	6	6	8	na	na	na	na
Staff qualifications and professional development[a]	16	3	6	8	8	16	3	6	8	8	16	na	na	na	na
Additional points in any domain		1	2	6	10		1	2	6	10		na	na	na	na
Minimum requirement for rating	50	16	26	38	42	50	16 points total and minimum points in at least two domains	26 points total and minimum points in at least three domains	38 points total and minimum points in at least four domains	42 points total and minimum points in all five domains	50	16 points total in any domain	26 points total in any domain	38 points total in any domain	42 points total in any domain
Program Quality Assessment (PQA) score		—	—	≥ 3.5	≥ 4.5		—	—	≥ 3.5	≥ 4.5		—	—	≥ 3.5	≥ 4.5

— is not available because the PQA is conducted only for programs with a self-assessment rating of level 4 or 5.

na is not applicable because the total score approach does not have domain score requirements.

a. Under the version 1.0 approach programs must have at least one point from each of three subdomains.

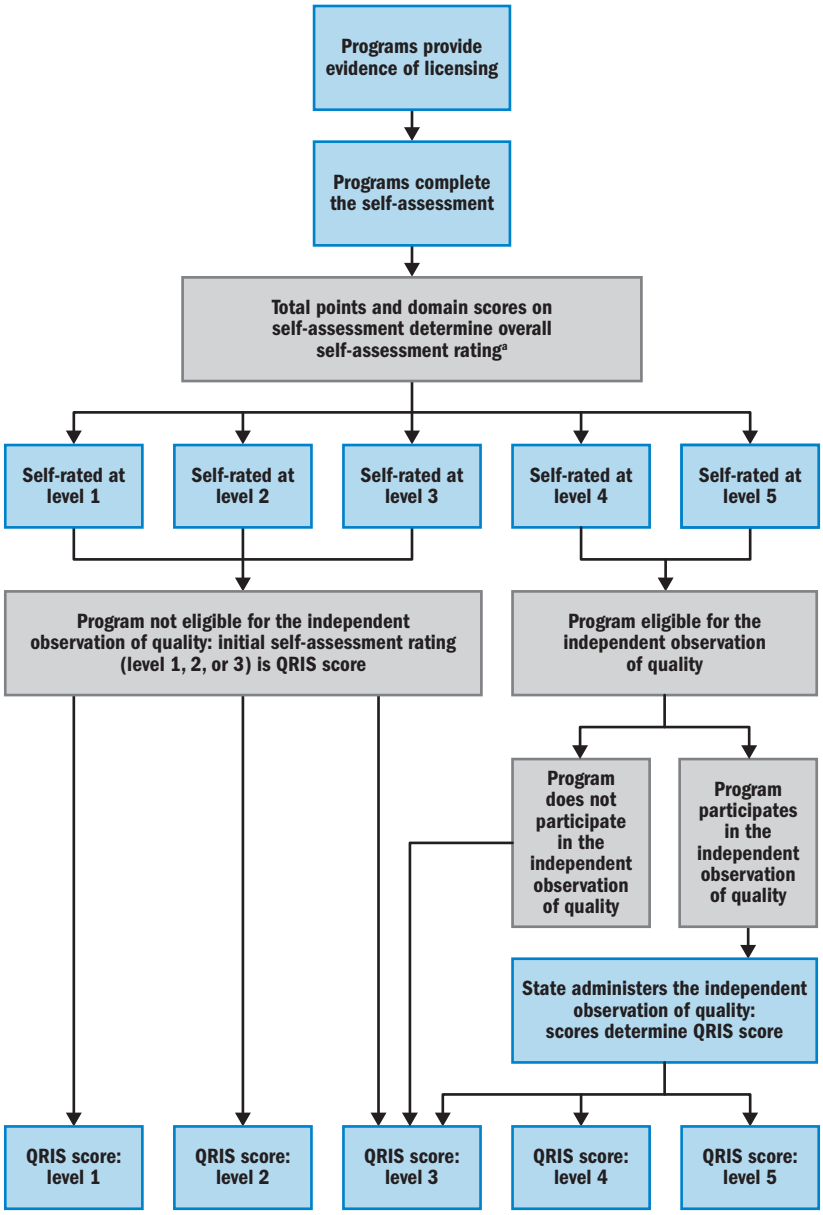
Source: Versions 1.0 and 2.0, materials shared by the Michigan Department of Education's Office of Great Start; total score approach, developed by authors.

At this stage programs with a self-assessment rating of 1, 2, or 3 have completed the steps required for a QRIS score. Programs with a self-assessment rating of 4 or 5 may voluntarily participate in the independent observation of quality and need to receive a minimum Program Quality Assessment score to receive a QRIS score of level 4 or 5. If a program has a self-assessment rating of level 4 or 5 but has not yet participated in the independent observation of quality, it does not have a QRIS score, and no information is reported until the independent observation is completed. Once the independent observation is completed, the minimum required points on form A of the Program Quality Assessment determine a program's QRIS score of level 3, 4, or 5 (see table A1 for cutscores). The rating on the independent observation overrides the self-assessment rating in determining the QRIS score, though programs must have high ratings on both to receive a QRIS score of 4 or 5. Programs that have a self-assessment rating of level 4 or 5 but have a score below 3.5 on the Program Quality Assessment receive a QRIS score of 3. Programs that have a self-assessment rating of level 5 but have a score above 3.5 and below 4.5 on the Program Quality Assessment receive a QRIS score of 4 (see table A1 for the cutscores for each dimension and rating level). Figure A2 illustrates the process of determining program ratings under the version 1.0 approach of Great Start to Quality.

Michigan recently changed the scoring structure of the self-assessment rating used in Great Start to Quality

Michigan developed the rating approach for Great Start to Quality over several years, launching the system statewide in 2012 with more than 2,000 programs participating by the end of the year. In January 2013 the state decided to revise its approach to calculating ratings as implementation of Great Start to Quality expanded. These changes went into effect in June 2013. Under the initial approach (version 1.0), the requirements for self-assessment ratings were based on both the total number of points and the number of points in each domain of the Self-Assessment Survey. The revised approach (version 2.0) continues to use the same self-assessment instrument but applies different cutscores on the self-assessment and changes the requirements for earning points on the staff qualifications subdomain (see table A1). Version 2.0 still requires a minimum number of points for each domain of the self-assessment but reduces the number of domains in which programs must meet those minimum required points for a rating of levels 2–4.

Figure A2. Steps to determine QRIS scores under the version 1.0 approach of Michigan's Great Start to Quality



QRIS is quality rating and improvement system.

a. The state verified the Self-Assessment Survey scores or a subsample of programs in the first year of implementation, based on a comprehensive review of their scores and documentation. For these programs the state-verified score replaced the initial self-assessment rating. See appendix E for more detail.

Note: Blue boxes refer to actions in the quality rating and improvement system (QRIS); gray boxes refer to decisions made by the state or program in the QRIS process.

Source: Documents shared with authors by Michigan's core QRIS team, including members of the Regional Educational Laboratory Midwest Early Childhood Education Research Alliance.

Appendix B. Description of quality rating improvement systems implemented in Regional Educational Laboratory Midwest Region states

Table B1 briefly describes the quality rating and improvement systems (QRISs) in the Regional Educational Laboratory (REL) Midwest Region states. Like states across the country, each state in the REL Midwest Region has a varied approach to quantifying a high-quality early childhood education program, and each state's QRIS measures different combinations of quality domains (Caronongan et al., 2011).

Table B1. Quality rating and improvement systems, by Regional Educational Laboratory Midwest Region state and year implemented

State	Name of quality rating and improvement system	Year implemented	Current scope of system	Number of rating levels	Approach to calculating ratings	Domains of quality measured										
						Accreditation	Administration and management	Child assessment	Community involvement	Curriculum	Environment	Family partnership	Health and safety	Licensing	Ratio/group size	Staff qualifications
Illinois	Quality Counts	2007	Statewide	4	Building block	●	●	●	●	●	●	●		●	●	●
Indiana	Paths to QUALITY	2001	Statewide	4	Building block	●	●	●	●	●	●	●		●	●	●
Iowa	Iowa Quality Rating System (or Iowa Child Care Quality System)	2006	Statewide	5	Hybrid	●	●				●	●	●	●	●	●
Michigan	Great Start to Quality	2011 pilot 2012 statewide	Statewide	5	Initially point, hybrid as of January 2013		●	●	●	●	●	●	●	●	●	●
Minnesota	Parent Aware	2007 pilot[a] 2012 county 2015 statewide	County	4	Hybrid	●		●		●	●	●	●	●		●
Ohio	Step Up to Quality	2006	Statewide	5	Hybrid	●	●	●	●	●	●	●		●	●	●
Wisconsin	YoungStar	2010	Statewide	5	Building bock	●	●	●		●	●		●	●	●	●

a. The five pilot areas in Minnesota were St. Paul, Minneapolis, Wayzata, Blue Earth, and Nicollet counties.

Source: Tout et al. (2010) and data collected via interview by the Regional Educational Laboratory Midwest Early Childhood Education Research Alliance.

Appendix C. Additional information on instruments

This appendix includes information on the instruments included under Great Start to Quality.

Michigan's Early Childhood Investment Corporation and Michigan's Great Start Collaborative maintain a database of all quality rating and improvement system (QRIS) data using MOSAIC software for all rated programs. This study relied on three data sources collected by Great Start to Quality and stored in the MOSAIC database: the Self-Assessment Survey, the Program Quality Assessment, and QRIS scores. Table C1 presents all variable names, data types, ranges, data sources, and the research questions they address.

Program Quality Assessment

The Program Quality Assessment measures seven domains of early childhood education quality: four are based on classroom observation (form A) and relate primarily to process quality, and three are based on interviews with teachers or directors (form B) and relate primarily to structural quality. Each domain has between 5 and 13 indicators of quality (tables C2 and C3). Each indicator is scored on a five-point scale, and indicator scores are averaged to calculate an overall score ranging from 1 to 5. Because Michigan uses form A to measure process quality (and the Self-Assessment Survey instrument is based on form B), Program Quality Assessment scores used in Great Start to Quality are an average of the items on form A only.

Table C1. Variables from Michigan's Early Childhood Investment Corporation and Michigan's Great Start Collaborative database used in the study

Variable name	Data type	Range	Data source	Research questions addressed
Self-assessment rating	Ordinal	1–5	SAS	1, 2, 3, 4
Self-assessment total score	Continuous	0–50	SAS	1, 2, 3, 4
Administration and management domain score	Continuous	0–6	SAS	2, 4
Curriculum domain score	Continuous	0–12 or 0–14[a]	SAS	2, 4
Environment domain score	Continuous	0–8	SAS	2, 4
Family and community partnerships domain score	Continuous	0–8	SAS	2, 4
Staff qualifications and professional development domain score	Continuous	0–16	SAS	2, 4
Additional bonus points in any domain	Continuous	0–18	SAS	2, 4
QRIS score	Ordinal	1–5	QRIS	1, 4
Independent observation of quality	Continuous	1.0–5.0	PQA	3, 4

SAS is Self-Assessment Survey. QRIS is quality rating improvement system. PQA is Program Quality Assessment.

a. Range was 0–12 for home-based family care early childhood education programs and 0–14 for center-based programs.

Source: Materials shared by the Michigan Department of Education's Office of Great Start.

Table C2. Indicators related to process quality on form A of the Program Quality Assessment, by domain

Adult–child interaction (13 indicators)	Curriculum planning and assessment (5 indicators)	Daily routine (12 indicators)	Learning environment (9 indicators)
Meeting basic physical needs	Curriculum model	Consistent daily routine	Safe and healthy environment
Handling separation from home	Team teaching	Parts of the day	Defined interest areas
Warm and caring atmosphere	Comprehensive child records	Appropriate time for each part of day	Logically located interest areas
Support for child communication	Anecdotal note taking by staff	Time for child planning	Outdoor space, equipment, materials
Support for non-English speakers	Use of child observation measure	Time for child-initiated activities	Organization and labeling of materials
Adults as partners in play		Time for child recall	Varied and open-ended materials
Encouragement of child initiative		Small-group time	Plentiful materials
Support for child learning at group times		Large-group time	Diversity-related materials
Opportunities for child exploration		Choices during transition times	Displays of child-initiated work
Acknowledgement of child efforts		Cleanup time with reasonable choices	
Encouragement for peer interactions		Snack or meal time	
Independent problem solving		Outside time	
Conflict resolution			

Source: High/Scope Educational Research Foundation, 2003.

Table C3. Indicators related to structural quality on form B of the Program Quality Assessment, by domain

Parent involvement and family services (10 indicators)	Program management (7 indicators)	Staff qualifications and staff development (7 indicators)
Opportunities for involvement	Program licensed	Program director background
Parents on policymaking committees	Continuity in instructional staff	Instructional staff background
Parent participation in child activities	Program assessment	Support staff orientation and supervision
Sharing of curriculum information	Recruitment and enrollment plan	Ongoing professional development
Staff–parent informal interactions	Operating policies and procedures	Inservice training content and methods
Extending learning at home	Accessibility for those with disabilities	Observation and feedback
Formal meetings with parents	Adequacy of program funding	Professional organization affiliation
Diagnostic/special education services		
Service referrals as needed		
Transition to kindergarten		

Source: High/Scope Educational Research Foundation, 2003.

The Program Quality Assessment is a valid and reliable measure of quality in early childhood education programs, although validation information is available only for the instrument as a whole (forms A and B combined). In a validity sample of 49 independent observations, Cronbach's alpha coefficient is .952, suggesting high internal consistency (High/Scope Educational Research Foundation, 2003). The Program Quality Assessment is significantly correlated overall with the Early Childhood Environment Rating Scale (r = .86; Harms & Clifford, 1980) and with the Arnett Global Rating Scale (r = .48; Arnett, 1989), suggesting moderate to high validity (High/Scope Educational Research Foundation, 2003). The associations were strongest in the subscales most similar in the instruments. For example, Program Quality Assessment learning environment and Early Childhood Environment Rating Scale furnishings are correlated at r = .73, and Program Quality Assessment adult–child interaction and Arnett Global Rating Scale sensitivity are correlated at r = .77 (High/Scope Educational Research Foundation, 2003).

Comparison of the Self-Assessment Survey and form A of the Program Quality Assessment

The Self-Assessment Survey, which was developed to mirror form B of the Program Quality Assessment, measures primarily structural quality, and form A of the Program Quality Assessment measures primarily process quality (see Lambert, Abbott-Shim, & Sibley, 2005, and Mashburn, 2008, for a description of structural and process quality). Table C4 compares the domains and subdomains measured by the Self-Assessment Survey and form A of the Program Quality Assessment and provides detailed information on how each instrument is scored and used in Great Start to Quality.

Although two domains on the Self-Assessment Survey and form A of the Program Quality Assessment appear to overlap (curriculum and environment), on further inspection the subdomains and indicators within each domain score are different, and thus the domains measure different content (tables C5 and C6).

Table C4. Characteristics of the self-assessment rating and independent observation components under Michigan's Great Start to Quality

Characteristic	Self assessment rating component		Independent observation component	
Instrument	Self-Assessment Survey		Form A of the Program Quality Assessment (PQA)	
Developer	Great Start to Quality staff, based primarily on High/Scope Educational Research Foundation's PQA form B		High/Scope Educational Research Foundation	
Use of the rating component	All programs participating in Great Start to Quality		Programs with self-assessment rating of level 4 or 5	
Assessor	Program staff complete instrument and upload results and documentation (state reviewed self-assessment ratings and documentation for some programs)		Independent observer hired by the state	
Type of quality measured	Mostly structural quality		Mostly process quality	
Domains and subdomains measured by instrument	**Domain**	**Subdomains**	**Domain**	**Subdomains**
	Administration and management	Administration and management (such as benefits, written policies, and staff evaluation)	Adult–child interaction	Teacher sensitivity, encouragement, and support for communication, learning, independence, and conflict resolution
	Curriculum	Curriculum (such as curriculum, daily routine, and cultural competence) Screening and assessment Consistent caregiving	Curriculum planning and assessment	Curriculum model, team teaching, record-keeping, and child observations
	Environment	Physical environment (such as safety) Staff ratios Health environment (such as outdoor time, nutrition plan, health records, and practices)	Daily routine	Consistency of daily routine, planning, and use of small and large groups; child initiation and choices; transitions; outdoor time; and meals
	Family and community partnerships	Family partnerships (such as communication with families and parent opportunities) Community partnerships (such as links to service agencies and transition supports)	Learning environment	Safety, health, space, materials, and classroom setup
	Staff qualifications and professional development	Administrator qualifications Lead teacher qualifications Assistant teacher qualifications Professional development		
Instrument scoring	0–50 points	1–2 points earned for each standard within five domains Domain scores summed for the overall score	1–5 points	Items scored 1–5 based on item criteria Item scores averaged for the overall score
Score requirements for each rating level	Level 1	Any Self-Assessment Survey score that does not qualify for level 2 or higher	Level 1	—
	Level 2	16 of 50 total and minimum domain scores	Level 2	—
	Level 3	26 of 50 total and minimum domain scores	Level 3	Score under 3.5, but PQA not required
	Level 4	38 of 50 total and minimum domain scores	Level 4	Score at or above 3.5 and less than 4.5
	Level 5	42 of 50 total and minimum domain scores	Level 5	Score at or above 4.5

— is not applicable because the PQA is conducted only for programs with a self-assessment rating of 4 or 5.

Source: Michigan Department of Education's Office of Great Start SAS instrument and PQA instrument.

Table C5. Comparison of curriculum domain, subdomains, and indicators under Michigan's Self-Assessment Survey and form A of the Program Quality Assessment

Self Assessment Survey			Form A of the Program Quality Assessment		
Domain	**Subdomain**	**Indicator**	**Domain**	**Indicator**	**Sample criteria used by observers to rate indicators**
Curriculum and instruction	Curriculum	A statement of educational and developmental priorities for the children.	Curriculum planning and assessment	Staff uses a comprehensive and documented curriculum model or educational approach to guide teaching practices.	Use of curriculum or educational approach; identifiable, documented model or approach; theory- or research-based teaching; and written curriculum statement providing rationale for practices and child development goals.
		A routine daily schedule that is predictable yet flexible, includes time for transition, includes indoor and outdoor activities, and is responsive to each child's need to be active or resting.		Staff uses a team teaching model and share responsibilities for planning and implementing program activities.	Use of team teaching model for planning and implementing activities, regularity of planning sessions, division of planning responsibilities, and involvement of assistants and aides.
	Screening and assessment	Uses an approved child assessment tool at least two times a year.		Staff regularly uses a child observation measure of proven reliability and validity to assess children's developmental progress.	Reliability of assessing children's developmental progress, use of child observation measures, and use of child observation measure in assessing developmental progress.
		Staff discusses anecdotal notes and observations as a basis for working with and teaching each child.			

Source: Materials shared by the Michigan Department of Education's Office of Great Start.

Table C6. Comparison of environment domain, subdomains, and indicators under Michigan's Self-Assessment Survey and form A of the Program Quality Assessment

Self Assessment Survey			Form A of the Program Quality Assessment		
Domain	**Subdomain**	**Indicator**	**Domain**	**Indicator**	**Sample criteria used by observers to rate indicators**
Environment	Physical environment	Physical environment. (Program is in a physical location that is free of environmental risks, such as lead, mercury, asbestos and indoor air pollutants.)	Learning environment	The classroom provides a safe and healthy environment for children.	Division of space, interest areas are clear and named, presence or lack of safety and health hazards, adequate lighting, temperature and ventilation, storage of nonprogram materials, presence of first-aid kit and evacuation plan.
				The space is divided into interest areas (for example, building or block area, house area, art area, toy area, book area, sand and water area) that address basic aspects of children's play and development.	Division of space, interest areas are clear, referenced by name, names are understood by children.
	Health environment	30 minutes of every three hours dedicated to active outdoor time, with appropriate indoor physical activities available when weather prohibits outdoor play.		An outdoor play area (at or near the program site) has adequate space, equipment, and materials to support various types of play. (Where extreme weather conditions or safety considerations prevent the regular use of outdoor play space, a large and open indoor space, such as a gymnasium, may be used as a substitute.)	Presence, size, use, and materials or equipment of outdoor play area, lack of health hazards in outdoor play area.
		A nutritional plan reviewed by a dietician or nutritionist.			

Source: Materials shared by the Michigan Department of Education's Office of Great Start.

Quality rating and improvement system scores

This study also used the QRIS scores for each program, as derived by Michigan from the Self-Assessment Survey and form A of the Program Quality Assessment. Each program had a QRIS score that ranged from level 1 to level 5, where level 1 indicates lower quality and level 5 indicates higher quality.

Appendix D. Analysis methods and detailed results for correlational analyses

This appendix includes additional details on the analysis methods and results used to address research question 2.

Analysis methods

The study team used a regression analysis framework to examine the strength of the relationships between each domain score on the Self-Assessment Survey instrument and the overall self-assessment ratings based on that instrument. The overall self-assessment ratings have hierarchical, discrete values ranging from 1 (lowest) to 5 (highest), so the regression analysis method used must account for the ordinal nature of the dependent variable. The appropriate regression technique depends on whether the magnitude of the difference in quality is the same for each change in rating level (for example, whether the difference between a level 1 and a level 2 rating is equivalent in "difficulty" to the difference between a level 2 and a level 3 rating, between a level 3 and a level 4 rating, and between a level 4 and a level 5 rating), indicated by the Brant test of the parallel regression assumption.

This study's approach was to use ordinal logistic regression if the Brant test had a nonsignificant result, indicating that it is appropriate to fit parallel regression lines with different intercepts but the same slope for all levels (Long & Freese, 2006), and to use general ordered logistic regression if the Brant test had a statistically significant result, indicating that both the intercept and the slope of the regression line should differ for each level (Fu, 1998; Williams, 2006). The Wald test was used to subsequently test whether each change in level was equivalent to the one immediately preceding it. A separate set of coefficients was estimated for each change in level with a significant result on the Wald test. These analyses included the Self-Assessment Survey scores for all 2,390 rated programs, including those that had and had not received an independent observation using form A of the Program Quality Assessment.

Detailed results

The Brant test results indicated that the proportional odds assumption was violated for the logistic regression analysis (table D1). Ordinal logistic regression would thus not be appropriate for the analysis. Instead, the study team used general ordered logistic regression as estimated by the gologit command in Stata (Fu, 1998) because it relaxes the assumption of parallel odds. Wald tests of the proportional odds assumption indicated that separate slope coefficients should be estimated for the move from level 1 to level 2, the move from level 2 to level 3, and the move from level 3 to level 4 but that a separate coefficient was not needed for the move from level 4 to level 5. The Wald test indicated that the coefficient for the move from level 3 to level 4 should not differ from the coefficient for the move from level 4 to level 5, so levels 4 and 5 were combined for the model estimation. Table D2 shows the full model results, which are described in the main body of the report.

Table D1. Brant test of parallel regression assumption

Brant test of parallel regression assumption	χ^2 statistic	*p* value
All domains	359.27	< .0001
Administration and management domain	100.45	< .0001
Curriculum domain	85.29	< .0001
Environment domain	113.90	< .0001
Family and community partnerships domain	46.27	< .0001
Staff qualifications and professional development domain	16.29	.001
Wald test of proportional odds assumption		
Equivalent coefficients for level 1 to level 2 cutoff and level 2 to level 3 cutoff	117.67	< .0001
Equivalent coefficients for level 2 to level 3 cutoff and level 3 to level 4 cutoff	75.71	< .0001
Equivalent coefficients for level 3 to level 4 cutoff and level 4 to level 5 cutoff	0.00	1.000

n = 2,390.

Note: Data are as of January 16, 2013.

Source: Authors' calculations based on Brandt tests of parallel regression assumption using data provided by the Michigan Department of Education's Office of Great Start on programs that had completed a Self-Assessment Survey.

Table D2. Generalized ordered logistic regression testing the relationship between each domain score and the overall Self-Assessment Survey rating under Michigan's Great Start to Quality

Parameter and domain	Point estimate (β)	Standard error	*p* value	Odds ratio	95 percent confidence interval	
					Minimum	Maximum
Likelihood of moving from SAS rating level 1 to level 2						
Administration and management	0.13	0.06	.026	1.14	0.02	0.24
Curriculum	0.21	0.03	< .0001	1.23	0.15	0.26
Environment	0.01	0.05	.807	1.01	–0.09	0.11
Family and community partnerships	0.30	0.04	< .0001	1.35	0.22	0.39
Staff qualifications and professional development	0.48	0.03	< .0001	1.62	0.43	0.54
Likelihood of moving from SAS rating level 2 to level 3						
Administration and management	0.72	0.07	< .0001	2.05	0.58	0.86
Curriculum	0.32	0.03	< .0001	1.38	0.26	0.38
Environment	0.08	0.05	.125	1.08	–0.02	0.18
Family and community partnerships	0.33	0.04	< .0001	1.39	0.24	0.42
Staff qualifications and professional development	0.55	0.03	< .0001	1.74	0.49	0.61
Likelihood of moving from SAS rating level 3 to level 4 or 5						
Administration and management	0.58	0.12	< .0001	1.79	0.35	0.81
Curriculum	0.66	0.05	< .0001	1.93	0.56	0.75
Environment	0.86	0.09	< .0001	2.37	0.70	1.03
Family and community partnerships	0.84	0.08	< .0001	2.32	0.69	0.99
Staff qualifications and professional development	0.80	0.05	< .0001	2.22	0.70	0.89

SAS is Self-Assessment Survey.

n = 2,390; model χ^2 = 4,132.12; *p* < .0001

Note: Data are as of January 16, 2013.

Source: Authors' calculations based on generalized ordered logistic regression using data provided by the Michigan Department of Education's Office of Great Start on programs that had completed an SAS.

Table D3 shows the odds ratio for each domain under several alternative approaches to specifying the generalized ordered logistic regression model, in addition to the main analysis model with levels 4 and 5 combined. Levels 4 and 5 were combined in response to a nonsignificant Wald test for comparing the increase from level 3 to level 4 and level 4 to level 5. However, the Wald test may have been nonsignificant in part because of model convergence problems due to multicolinearity—because the five domain scores also perfectly predicted the self-assessment rating in the model with the level 3 to level 4 coefficient estimated separately from the level 4 to level 5 coefficient. Table D3 shows alternative approaches to estimating the model; in these alternative models, the Wald test indicated that separate coefficients should be estimated for all the rating level increases, including level 3 to level 4 and level 4 to level 5. The left-most data column includes estimates for each domain run separately, with no other predictors in the model. The five middle data columns show the results when the model is run with just four of the five domains, eliminating the problem of perfect prediction. The table shows that the model results can depend on the estimation approach, with especially variable results when one domain is dropped from the model. The four-level model was chosen for the study because all five domains are included and the estimates control for all other domain scores.

Table D3. Odds ratios from alternative generalized ordered logistic regression models indicate that the relationship between each domain and the self-assessment rating varies by how the model is specified

Parameter and domain	Each domain run separately	Five level general ordered logistic regression					Four level general ordered logistic regression, all five domains in model
		Without administrative domain	Without curriculum domain	Without environment domain	Without family domain	Without staff domain	
Likelihood of moving from SAS rating level 1 to level 2							
Administration and management	1.84***	—	1.23	1.14	1.23***	1.13***	1.14*
Curriculum	1.66***	1.26***	—	1.23***	1.32***	1.43***	1.23***
Environment	1.62***	1.00	1.08***	—	1.09	1.02	1.01
Family and community partnerships	1.95***	1.41***	1.61***	1.37***	—	1.34***	1.35***
Staff qualifications and professional development	1.84***	1.66***	1.80***	1.65***	1.62***	—	1.62***
Likelihood of moving from SAS rating level 2 to level 3							
Administration and management	2.35***	—	2.11*	2.01*	2.14***	1.60***	2.05***
Curriculum	1.78***	1.37***	—	1.36***	1.46***	1.51***	1.38***
Environment	1.74***	1.14*	1.17***	—	1.15*	1.07*	1.08
Family and community partnerships	2.10***	1.47***	1.60***	1.38***	—	1.37***	1.39***
Staff qualifications and professional development	1.76***	1.55***	1.77***	1.71***	1.72***	—	1.74***
Likelihood of moving from SAS rating level 3 to level 4 or from level 3 to levels 4 and 5[a]							
Administration and management	2.85***	—	2.27***	1.62***	1.86***	1.45***	1.78***
Curriculum	2.29***	1.87***	—	1.70***	1.92***	1.81***	1.93***
Environment	2.24***	2.15***	2.01***	—	2.16***	1.52***	2.37***
Family and community partnerships	2.83***	2.27***	2.47***	2.07***	—	1.73***	2.32***
Staff qualifications and professional development	1.88***	2.00***	2.00***	1.74***	1.82***	—	2.22***
Likelihood of moving from SAS rating level 4 to level 5							
Administration and management	3.49***	—	3.53***	5.95***	5.40***	2.65***	—
Curriculum	3.38***	10.31***	—	7.25***	8.27***	3.47***	—
Environment	2.70***	10.55***	3.29***	—	7.68***	2.99***	—
Family and community partnerships	3.91***	8.28***	4.44***	7.84***	—	3.65***	—
Staff qualifications and professional development	2.26***	7.54***	2.93***	6.07***	6.50***	—	—

SAS is Self-Assessment Survey.

— is not applicable.

* $p < .05$; ** $p < .01$; *** $p < .0001$.

$n = 2{,}390$, model $\chi^2 = 4132.12$, $p < .0001$.

Note: Data are as of January 16, 2013.

a. The odds ratio refers to the likelihood of moving from level 3 to level 4 for the five-level models and of moving from level 3 to level 4 or 5 for the four-level model.

Source: Authors' calculations based on alternative generalized ordered logistic regression using data provided by the Michigan Department of Education's Office of Great Start on programs that had completed an SAS.

Table D4 shows the nonparametric Spearman's rho correlations between the domain scores and the self-assessment ratings. As shown in the table, all domains are moderately to highly correlated with the self-assessment rating and with each other.

Table D4. Domain scores are moderately to highly correlated with the self-assessment ratings and with each other

Domain	Overall self assessment rating (level 1, 2, 3, 4, or 5)	Administration and management	Curriculum	Environment	Family and community partnerships	Staff qualifications and professional development
Administration and management	.51***					
Curriculum	.79***	.54***				
Environment	.49***	.43***	.50**			
Family and community partnerships	.69***	.53***	.73**	.51***		
Staff qualifications and professional development	.83***	45***	.71**	.41***	.59***	

* $p < .05$; ** $p < .01$; *** $p < .0001$.

Note: Data are as of January 16, 2013.

Source: Authors' calculations based on Spearman's rho correlation coefficients using data provided by the Michigan Department of Education's Office of Great Start on programs that had completed a Self-Assessment Survey.

Appendix E. Programs in the study sample

This appendix provides more information on the programs in the study sample. It describes four subsamples and provides descriptive statistics for each:

- Programs that participated in Great Start to Quality in any way.
- Programs that completed the self-assessment.
- Programs whose self-assessment rating was verified by the state.
- Programs eligible for the independent observation of quality.

All program information is as of January 16, 2013.

Programs that participated in Great Start to Quality in any way

Great Start to Quality is voluntary, and thus not all early childhood education programs in Michigan are rated by it. Some 3,941 programs (35.6 percent) in Michigan had participated. Licensed child care centers accounted for the largest share of participating programs (54.3 percent), followed by registered family child care homes (27.6 percent) and licensed group child care homes (18.1 percent; table E1). A majority of participating programs were preschool programs (51.3 percent). Participating programs had a higher mean license capacity (41.9) than all programs did (31.5).

Table E1. Descriptive statistics for all early childhood education programs and programs that participated in Michigan's Great Start to Quality

Statistic	All programs		Participating programs	
Characteristic	Number	Mean (standard deviation)	Number	Mean (standard deviation)
License capacity (program size)	11,081	31.5 (43.9)	3,941	41.9 (51.7)
Type[a]	**Number**	**Share of total (%)**	**Number**	**Share of total (%)**
Licensed child care center	4,424	40.0	2,140	54.3
Registered family child care home	4,439	40.0	1,089	27.6
Licensed group child care home	2,209	20.0	712	18.1
Education approach[a]	**Number**	**Share of total (%)**	**Number**	**Share of total (%)**
Preschool	3,124	28.2	2,021	51.3
Great Start Readiness Program	372	3.4	333	8.4
Religious	483	4.4	307	7.8
Head Start	320	2.9	298	7.6
National Association for the Education of Young Children	181	1.6	148	3.8
Montessori	153	1.4	93	2.4
Reggio-inspired	90	0.8	71	1.8
Early Head Start	43	0.4	41	1.0
Other	1,191	10.8	783	19.9

Note: Data are as of January 16, 2013.

a. Components do not sum to total and percentages do not sum to 100 because some programs are unclassified and some are classified in multiple categories.

Source: Descriptive demographic information provided by the Michigan Department of Education's Office of Great Start.

The regions most represented among participating programs were Southeast (13.9 percent), Oakland-Macomb (13.5 percent), and Central (13.1 percent; table E2). However, these regions are underrepresented when compared with their share among all programs, while the Northeast and Upper Peninsula regions are overrepresented.

Programs that completed the Self-Assessment Survey

Of the 3,941 programs that participated in Michigan's Great Start to Quality, 2,390 (60.6 percent) completed the Self-Assessment Survey. Of the 1,551 programs that started but did not complete the Self-Assessment Survey, 42.4 were registered family child care home programs (higher than their share among all programs), 32.8 percent were licensed child care centers (lower than their share among all programs), and 24.8 percent were licensed group child care homes (higher than their share among all programs). Preschool, Head Start, and Great Start Readiness programs accounted for the largest shares of programs that completed the Self-Assessment Survey (table E3).

Programs from the Central and Southeast regions made up the largest percentage of programs that did not complete the Self-Assessment Survey (table E4).

Programs whose Self-Assessment Survey rating was verified by the state

Of the 2,390 programs that completed the self-assessment, 470 had their rating and documentation verified by the state, and 446 of those had both the original self-assessment rating and the state-verified rating. Programs may have missing data if their review is currently in progress. Among these 446 programs, when the two ratings differed, self-assessment ratings tended to be higher than state-verified ratings (under the version 1.0

Table E2. Regional distribution of all early childhood education programs and programs that participated in Michigan's Great Start to Quality

	All programs[a] (*n* = 11,072)		Participating programs (*n* = 3,941)	
Region	Number	Share of total[b] (%)	Number	Share of total[b] (%)
Central	1,251	11.3	516	13.1
Eastern	836	7.5	303	7.7
Kent	982	8.8	252	6.4
Northeast	355	3.2	215	5.5
Northwest	513	4.6	221	5.6
Oakland-Macomb	1,687	15.2	532	13.5
Southeast	1,515	13.7	549	13.9
Southwest	1,309	11.8	386	9.8
Upper Peninsula	295	2.7	216	5.5
Wayne County	1,249	11.3	472	12.0
Western	1,080	9.8	279	7.1

a. Nine programs were missing geographic data in the state's data file and are thus excluded from the table.

b. Does not sum to 100 percent because of rounding.

Note: Data are as of January 16, 2013.

Source: Descriptive demographic information provided by the Michigan Department of Education's Office of Great Start.

Table E3. Descriptive statistics for programs that completed the Self-Assessment Survey under Michigan's Great Start to Quality and programs that began but did not complete it

Statistic	Completed Self Assessment Survey		Began but did not complete Self Assessment Survey	
Characteristic	**Number**	**Mean (standard deviation)**	**Number**	**Mean (standard deviation)**
License capacity (program size)	2,390	49.9 (54.4)	1,551	29.5 (44.6)
Type[a]	**Number**	**Share of total (%)**	**Number**	**Share of total (%)**
Licensed child care center	1,631	68.2	509	32.8
Registered family child care home	431	18.0	658	42.4
Licensed group child care home	328	13.7	384	24.8
Education approach[a]	**Number[a]**	**Share of total[a] (%)**	**Number[a]**	**Share of total[a] (%)**
Preschool	1,345	56.3	676	43.6
Great Start Readiness Program	301	12.6	32	2.1
Religious	178	7.4	129	8.3
Head Start	278	11.6	20	1.3
Montessori	63	2.6	30	1.9
Reggio-inspired	57	2.4	14	0.9
Early Head Start	38	1.6	3	0.2
Other	525	22.0	258	16.6
Accreditation	**Number**	**Percent**	**Number**	**Percent**
National Association for the Education of Young Children	122	5.1	26	1.7
National Association for Family Child Care	45	1.9	41	2.6

Note: Data are as of January 16, 2013.

a. Components do not sum to total and percentages do not sum to 100 because some programs are unclassified and some are classified in multiple categories.

Source: Descriptive demographic information provided by the Michigan Department of Education's Office of Great Start.

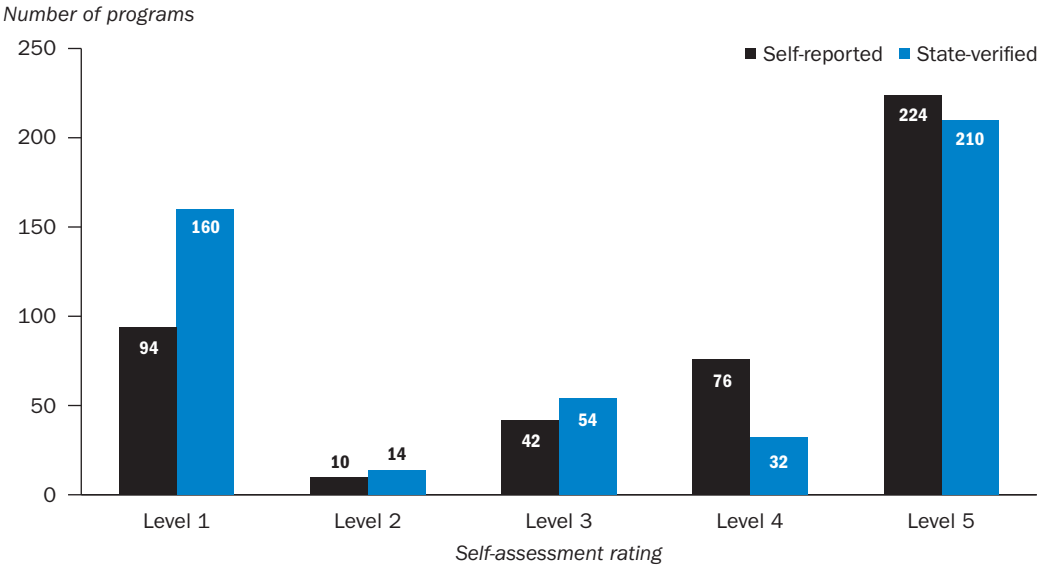
Table E4. Distribution of programs that completed the Self-Assessment Survey under Michigan's Great Start to Quality and programs that began but did not complete it, by region

Region	Completed Self Assessment Survey (*n* = 2,390)		Began but did not complete Self Assessment Survey (*n* = 1,551)	
	Number	Percent	Number	Percent
Central	256	10.7	260	16.8
Eastern	174	7.3	129	8.3
Kent	121	5.1	131	8.4
Northeast	147	6.2	68	4.4
Northwest	162	6.8	59	3.8
Oakland-Macomb	346	14.5	186	12.0
Southeast	306	12.8	243	15.7
Southwest	235	9.8	151	9.7
Upper Peninsula	162	6.8	54	3.5
Wayne County	340	14.2	132	8.5
Western	141	5.9	138	8.9

Note: Data are as of January 16, 2013.

Source: Descriptive demographic information provided by the Michigan Department of Education's Office of Great Start.

Figure E1. Self-assessment ratings and state-verified ratings differed for 40 percent of programs



n = 446.

Note: Data are as of January 16, 2013.

Source: Michigan Department of Education's Office of Great Start.

approach of Great Start to Quality; figure E1), though they were not always. The two ratings were the same for 59.4 percent of programs, state-verified ratings were lower than self-reported ratings for 24.7 percent of programs, and state-verified ratings were higher than the self-reported ratings for 15.9 percent of programs (table E5). Although the state

Table E5. Self-reported ratings on the Self-Assessment Survey and state-verified ratings under Michigan's Great Start to Quality version 1.0, by rating level

Self-reported rating	State-verified rating				
	Level 1 (*n* = 160)	Level 2 (*n* = 14)	Level 3 (*n* = 54)	Level 4 (*n* = 32)	Level 5 (*n* = 210)
Level 1 (*n* = 94)	**86**	3	4	1	0
Level 2 (*n* = 10)	4	**1**	2	2	1
Level 3 (*n* = 42)	11	4	**13**	6	8
Level 4 (*n* = 76)	14	3	6	**9**	44
Level 5 (*n* = 224)	29	3	22	14	**156**

n = 446.

Note: Bolding indicates that self-reported and state-verified ratings were the same. Data are as of January 16, 2013.

Source: Michigan Quality Rating Improvement Survey data on programs' scores on the Self-Assessment Survey (SAS) and state-verified SAS scores shared by the Michigan Department of Education's Office of Great Start.

verified ratings for programs that self-rated at all levels, those that self-rated at level 4 or 5 were verified more often than those that self-rated at level 1, 2, or 3.

Programs eligible for the independent observation of quality

Of the 2,390 programs that completed the self-assessment, those rated at level 4 or level 5 were eligible to participate in the independent observation of quality using form A of the Program Quality Assessment. Of the 1,049 eligible programs, 72 conducted an independent observation. The mean, median, and range of self-assessment ratings do not seem to differ between programs that conducted the independent observation and the full sample of eligible programs (table E6). However, only center-based care programs (no licensed family or home child care programs) were included among programs with the independent observation.

Table E6. Descriptive characteristics of programs with self-assessment ratings of levels 4 and 5 under Michigan's Great Start to Quality, by independent observation of quality status

Program characteristic or type	Self assessment rating of level 4		Self assessment rating of level 5	
	Did not conduct independent observation	Conducted independent observation	Did not conduct independent observation	Conducted independent observation
Characteristic				
Self-assessment total score				
Mean	39.56	39.89	46.13	46.17
Standard deviation	1.17	1.05	2.39	2.30
Median	40	40	46	46
Range	38–41	38–41	42–50	42–50
Type (percent of total)				
Licensed child care center	85.9	100.0	92.7	100.0
Licensed group child care home	6.3	0.0	4.5	0.0
Registered family child care home	7.8	0.0	2.8	0.0

Note: Data are as of January 16, 2013.

Source: Descriptive demographic information shared by the Michigan Department of Education's Office of Great Start.

Some locations were disproportionately represented among programs with an independent observation of quality (table E7). For example, Kalamazoo County accounted for 3.6 percent of eligible programs and Saginaw County for 2.2 percent, but each accounted for 15.3 percent of programs with an independent observation. Wayne County accounted for 12.9 percent of eligible programs but just 1.4 percent of programs with an independent observation. Because programs were not randomly sampled and representative of the statewide distribution of programs, generalizations from results of the analyses should be made with caution.

Table E7. Distribution of programs eligible for an independent observation of quality under Michigan's Great Start to Quality, by county and independent observation status

	All programs eligible for independent observation (*n* = 1,049)		Did not conduct independent observation (*n* = 977)		Conducted independent observation (*n* = 72)	
County	Number	Percent	Number	Percent	Number	Percent
Kalamazoo	38	3.6	27	2.8	11	15.3
Saginaw	23	2.2	12	1.2	11	15.3
Wayne	135	12.9	134	13.7	1	1.4
All other counties	853	81.3	804	82.3	49	68.1

Note: Data are as of January 16, 2013.

Source: Descriptive demographic information provided by the Michigan Department of Education's Office of Great Start.

Appendix F. Descriptive statistics for the version 1.0 and 2.0 approaches of Michigan's Great Start to Quality, by rating and domain

This appendix includes additional descriptive statistics for version 1.0 and version 2.0 of Michigan's quality rating and improvement system (QRIS), Great Start to Quality.

Self-assessment total scores and domain scores increased from level 1 to level 5, with some notable exceptions

Among the 1,413 programs with a QRIS score (not the 2,390 with a self-assessment rating), self-assessment total scores under the version 1.0 approach increased with the QRIS score (table F1). However, the amount of the increase was not consistent across rating levels. The difference between the mean total score was much smaller between levels 4 and 5 than between other levels. In general, mean domain scores increased with the QRIS score, but again the difference between levels 4 and 5 was small, null, or negative, except in the staff qualifications and professional development domain. And the mean administration and management domain score was higher for level 1 than level 2. Mean Program Quality Assessment scores also increased from level 3 to level 5.

Table F2 presents the distribution of QRIS scores under version 2.0. Again, self-assessment total scores and domain scores increased as the QRIS score increased. However, the largest increases are between levels 1 and 2, and the differences between levels 4 and 5 are minimal for total score and domain scores.

Figure F1 shows the overlap in Self-Assessment Survey total scores among programs rated at level 4 or level 5 that also had an independent observation under the version 1.0 approach. It also highlights the heterogeneity of the programs with a rating of level 3, with many outliers at both the high and low ends of the Self-Assessment Survey total score. This distribution is consistent with other QRISs across the country, where identifying and classifying programs as moderate in quality is more difficult than identifying low- or high-quality programs (Caronongan et al., 2011).

Under the version 2.0 approach, the distribution includes fewer outliers at level 3 and has clearer distinctions on the Self-Assessment Survey total score among QRIS scores of levels 3, 4, and 5. However, the new system also includes programs that are now considered outliers at level 4, with higher Self-Assessment Survey total scores than most other programs at level 4. This distribution suggests that Michigan's shift from the version 1.0 approach to the version 2.0 approach could mean that moderate-quality programs are categorized more accurately.

Descriptive statistics by domain show little variability for programs rated at levels 4 and 5 under both the version 1.0 and 2.0 approaches

Figures F2–F6 depict box-and-whisker plots for each Self-Assessment Survey domain under the version 1.0 and 2.0 approaches. The plots suggest a possible ceiling effect on some domains, where many programs achieved the highest possible domain score. For example, on the family and community partnership domain, all but four programs at level 4 or 5 scored the maximum (8 points; see figure F5). The pattern is similar for the administration and management, curriculum, and environment domains under both approaches. These scores could suggest that the Self-Assessment Survey needs further psychometric testing for use with early child care programs in Michigan.

Table F1. Mean total scores and domain scores on the self-assessment and independent observation of quality tend to increase with the QRIS score under the version 1.0 approach of Great Start to Quality

	QRIS score														
	Level 1 (*n* = 891)			Level 2 (*n* = 117)			Level 3 (*n* = 340)			Level 4 (*n* = 44)			Level 5 (*n* = 21)		
Domain	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Self-assessment total score	20.14	11.33	0–46	31.22	5.16	19–44	36.10	3.89	26–49	45.11	3.07	38–49	46.33	2.29	42–50
Administration and management	3.93	2.27	0–8	3.59	1.75	2–8	5.46	0.89	4–6	6	0.00	6–6	6	0.00	6–6
Curriculum	4.22	3.93	0–12	7.15	3.00	4–12	8.92	2.18	6–12	11.59	1.10	8–12	11.81	0.60	10–12
Environment	5.22	2.67	0–8	5.90	1.68	2–8	6.26	1.47	4–8	7.45	0.90	6–8	7.62	0.81	6–8
Family and community partnerships	3.61	2.95	0–8	5.76	1.60	4–8	6.26	1.59	4–8	7.86	0.51	6–8	7.81	0.60	6–8
Staff qualifications and professional development	3.16	3.42	0–16	8.83	3.03	3–16	9.20	2.53	6–16	12.2	2.16	8–15	13.1	2.16	9–16
	Level 1 (*n* = 0)			Level 2 (*n* = 0)			Level 3 (*n* = 7)			Level 4 (*n* = 44)			Level 5 (*n* = 21)		
Independent observation of quality rating (form A of the Program Quality Assessment)		na			na		3.24	0.21	2.9–3.4	4.06	0.24	3.5–4.4	4.73	0.15	4.5–5.0

QRIS is quality rating and improvement system.

na is not applicable because no programs were rated at level 1 or 2.

SD is standard deviation.

Note: Data are as of January 16, 2013.

Source: Michigan quality rating and improvement system data provided by the Michigan Department of Education's Office of Great Start.

Table F2. Mean total scores and domain scores on the self-assessment and independent observation of quality tend to increase with the QRIS score under the version 2.0 approach of Great Start to Quality

	QRIS score														
	Level 1 (*n* = 281)			Level 2 (*n* = 308)			Level 3 (*n* = 599)			Level 4 (*n* = 160)			Level 5 (*n* = 74)		
Domain	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Self-assessment total score	7.88	8.53	0–47	21.54	5.15	0–43	33.00	4.78	10–50	40.99	3.32	26–50	45.66	3.08	36–50
Administration and management	1.77	2.19	0–6	4.44	1.64	0–6	5.06	1.34	0–6	5.56	1.14	2–6	5.97	0.23	4–6
Curriculum	0.65	1.35	0–6	3.57	2.26	0–10	7.68	2.64	0–12	10.80	2.00	4–12	11.84	0.64	8–12
Environment	2.70	2.70	0–8	5.84	1.80	0–8	6.38	1.48	2–8	6.76	1.45	4–8	7.62	0.79	6–8
Family and community partnerships	0.78	1.35	0–6	3.54	2.12	0–8	5.99	1.84	0–8	6.95	1.63	2–8	7.89	0.46	6–8
Staff qualifications and professional development	0.65	1.68	0–12	3.21	2.95	0–13	7.03	3.32	0–16	9.81	3.19	4–16	12.66	2.05	8–16
	1 (*n* = 0)			2 (*n* = 0)			3 (*n* = 0)			4 (*n* = 9)			5 (*n* = 63)		
Independent observation of quality rating (form A of the Program Quality Assessment)		na			na			na		3.77	0.39	3.2–4.3	4.23	0.47	2.9–5.0

QRIS is quality rating and improvement system.

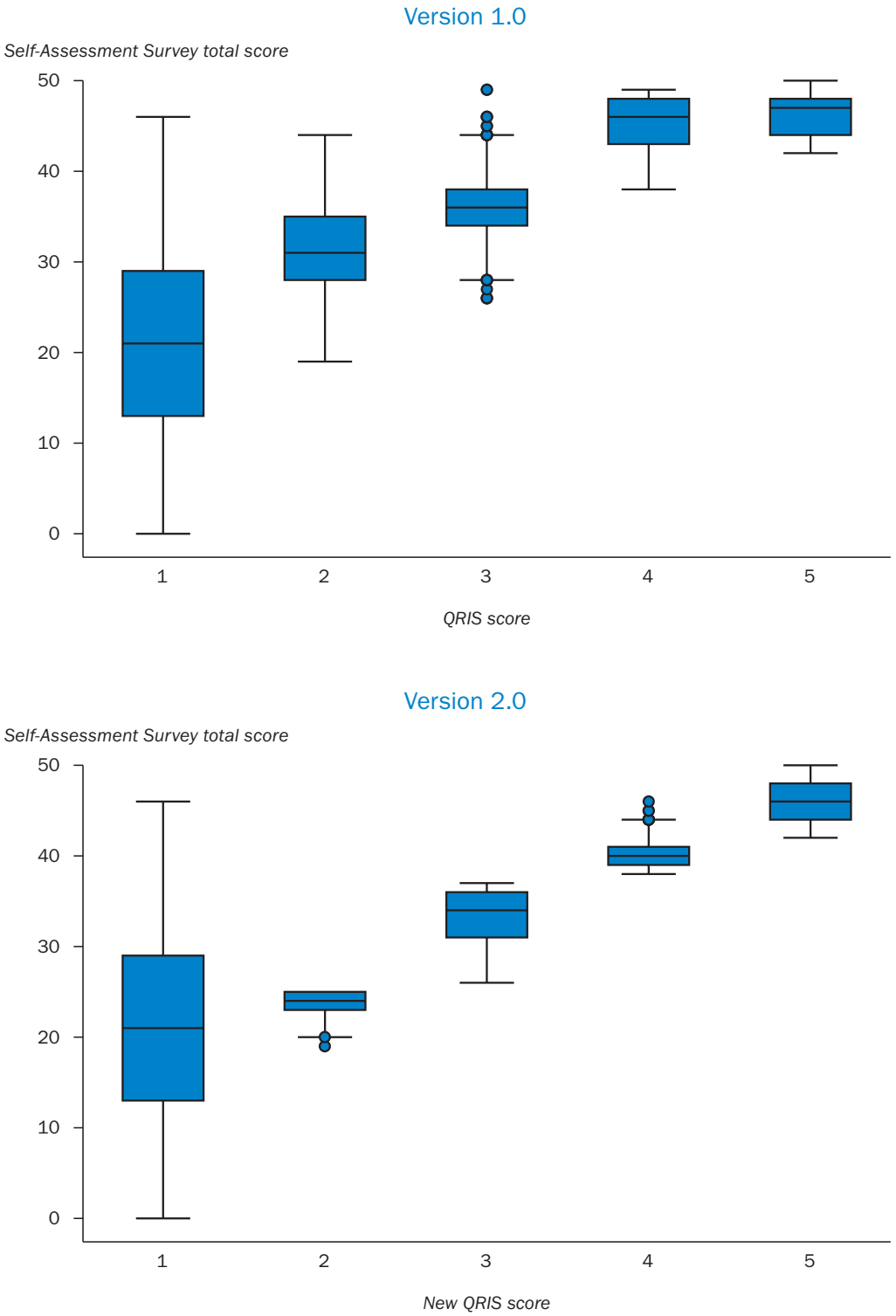
na is not applicable because no programs were rated at level 1, 2, or 3.

SD is standard deviation.

Note: Data are as of January 16, 2013.

Source: Authors' calculations using data on the version 1.0 approach provided by the Michigan Department of Education's Office of Great Start.

Figure F1. Box-and-whisker plots of Self-Assessment Survey total scores, by QRIS score, under the version 1.0 and 2.0 approaches of Michigan's Great Start to Quality



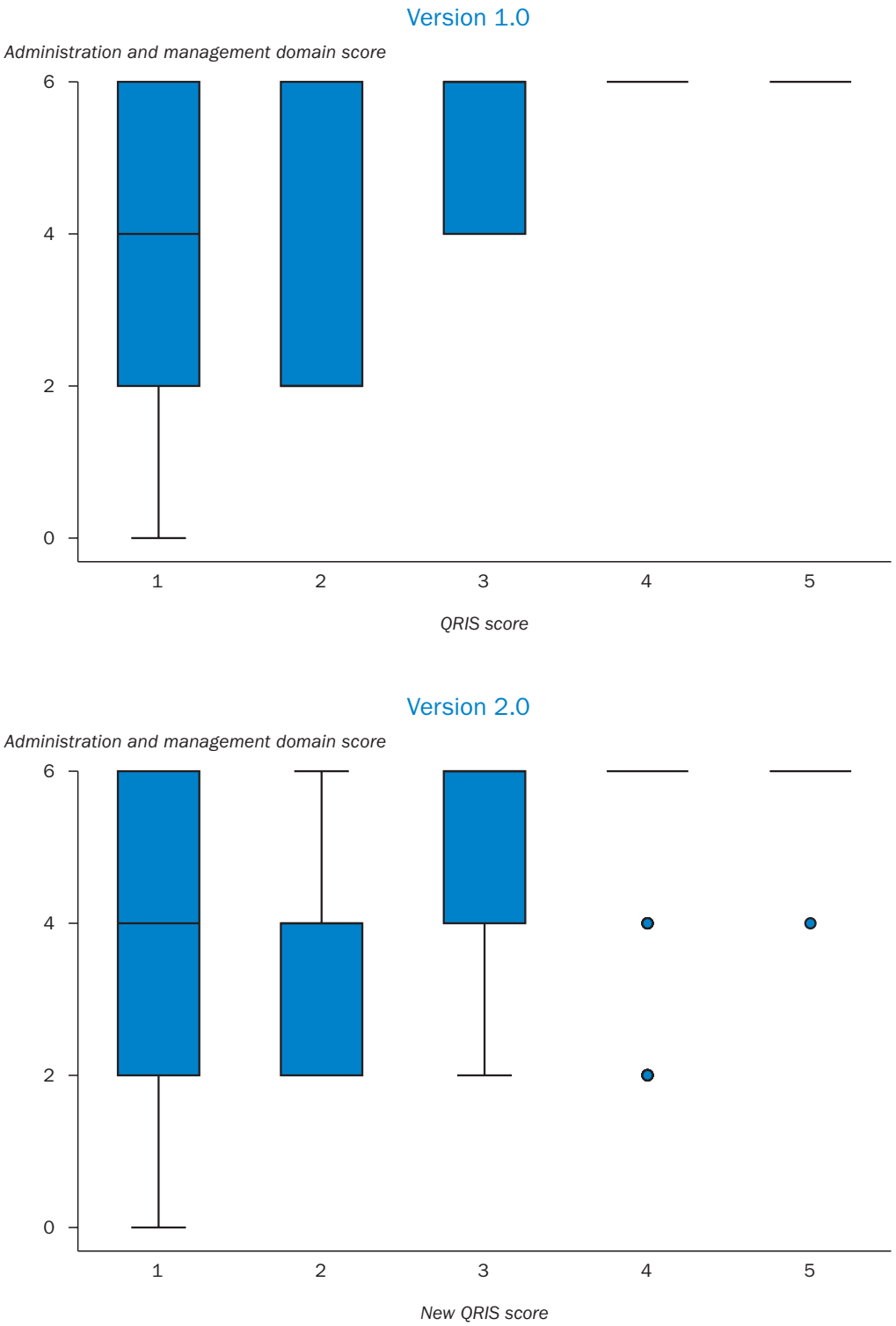
QRIS is quality rating and improvement system.

n = 1,413.

Note: The maximum value on the self-assessment survey score is 50. Circles represent data points that are considered outliers. Data are as of January 16, 2013.

Source: Version 1.0, data provided by the Michigan Department of Education's Office of Great Start; version 2.0, authors' calculations using those data.

Figure F2. Box-and-whisker plots for the administration and management domain of the Self-Assessment Survey, by QRIS score, under the version 1.0 and 2.0 approaches of Michigan's Great Start to Quality



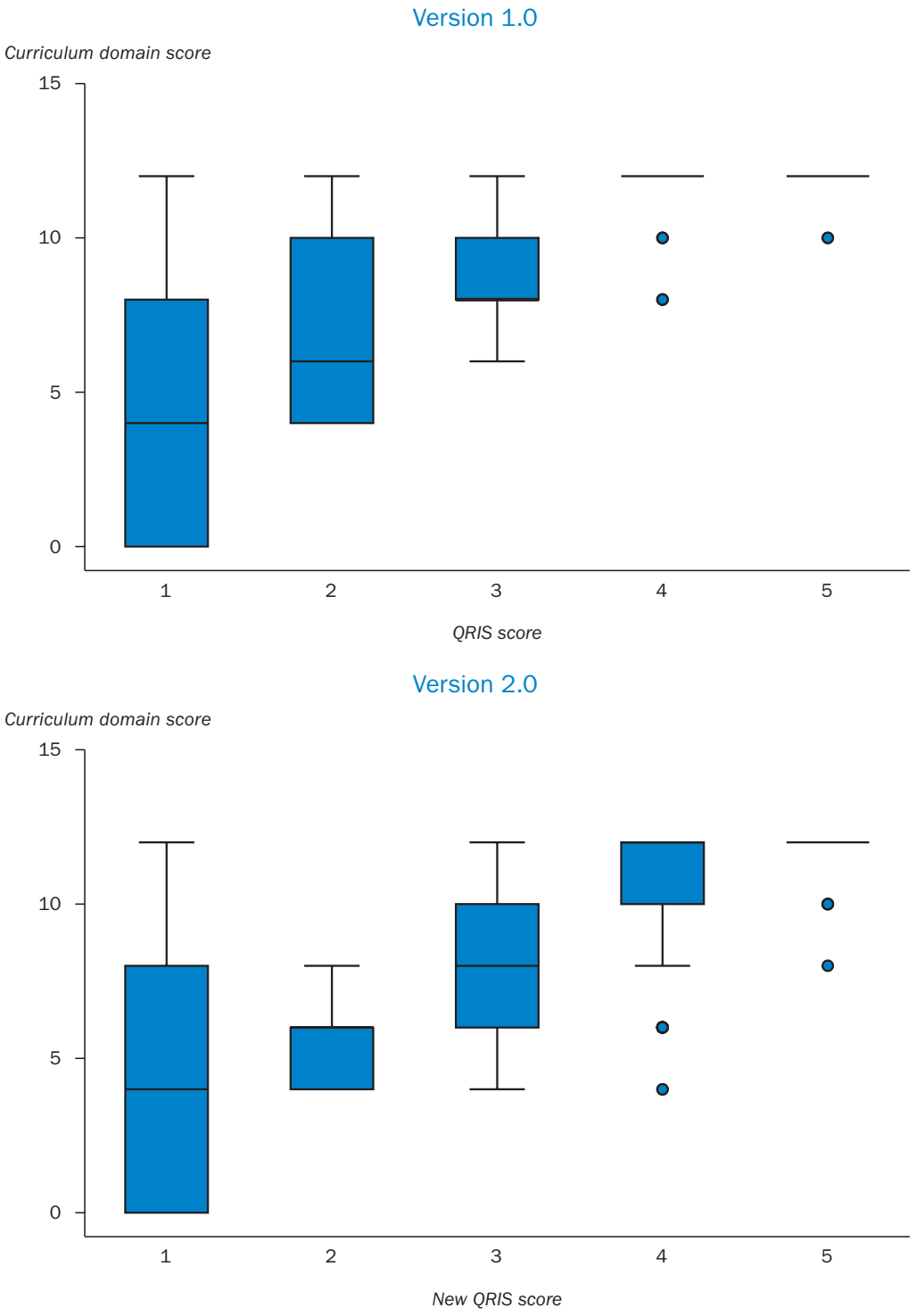
QRIS is quality rating and improvement system.

n = 1,413.

Note: The maximum value on the administration and management domain score is 6. Circles represent data points that are considered outliers. Data are as of January 16, 2013.

Source: Version 1.0, data provided by the Michigan Department of Education's Office of Great Start; version 2.0, authors' calculations using those data.

Figure F3. Box-and-whisker plots for the curriculum domain of the Self-Assessment Survey, by QRIS score, under the version 1.0 and 2.0 approaches of Michigan's Great Start to Quality



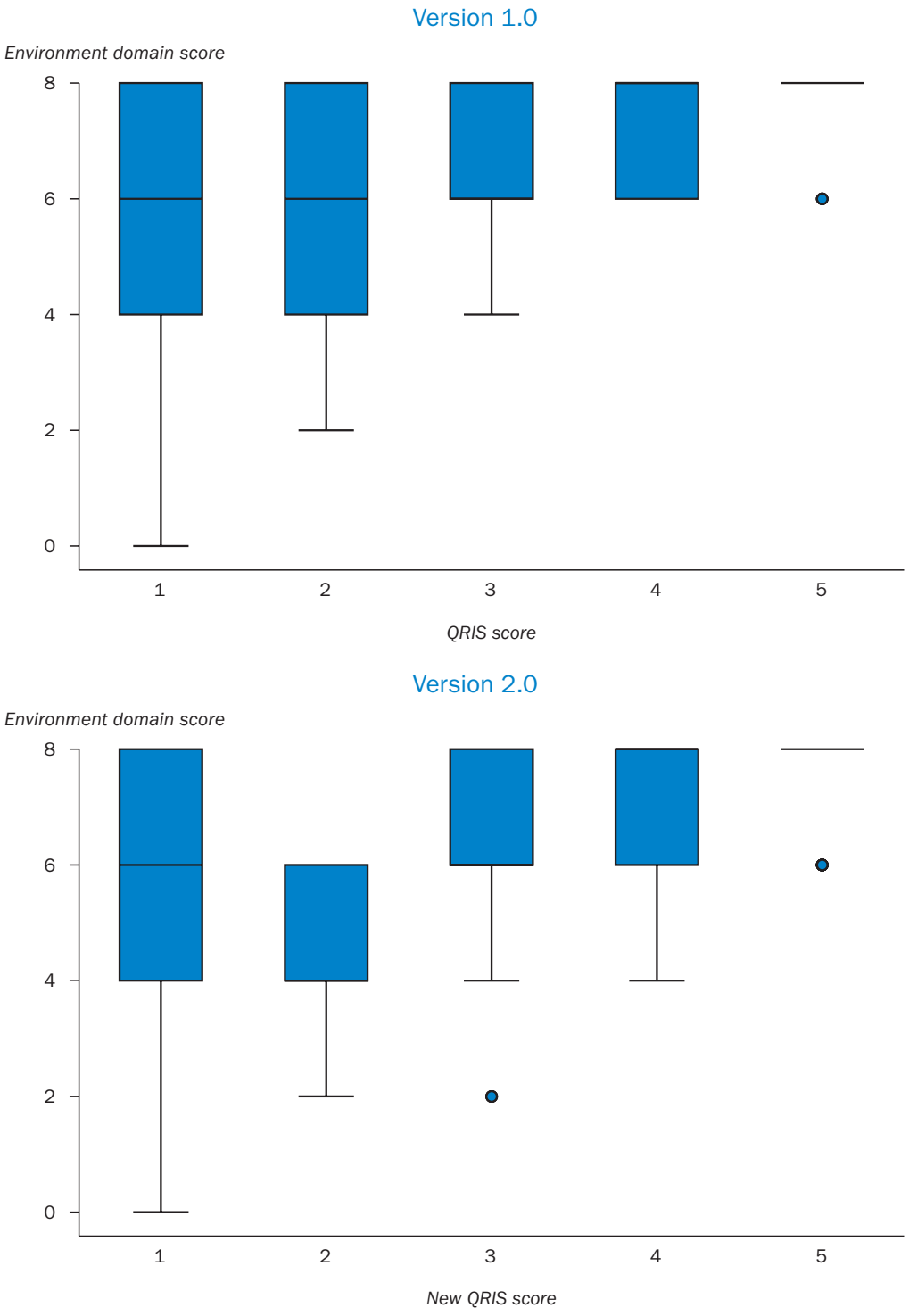
QRIS is quality rating and improvement system.

n = 1,413.

Note: The maximum value on the curriculum domain score is 12. Circles represent data points that are considered outliers. Data are as of January 16, 2013.

Source: Version 1.0, data provided by the Michigan Department of Education's Office of Great Start; version 2.0, authors' calculations using those data.

Figure F4. Box-and-whisker plots for the environment domain of the Self-Assessment Survey by QRIS score, under the version 1.0 and 2.0 approaches of Michigan's Great Start to Quality



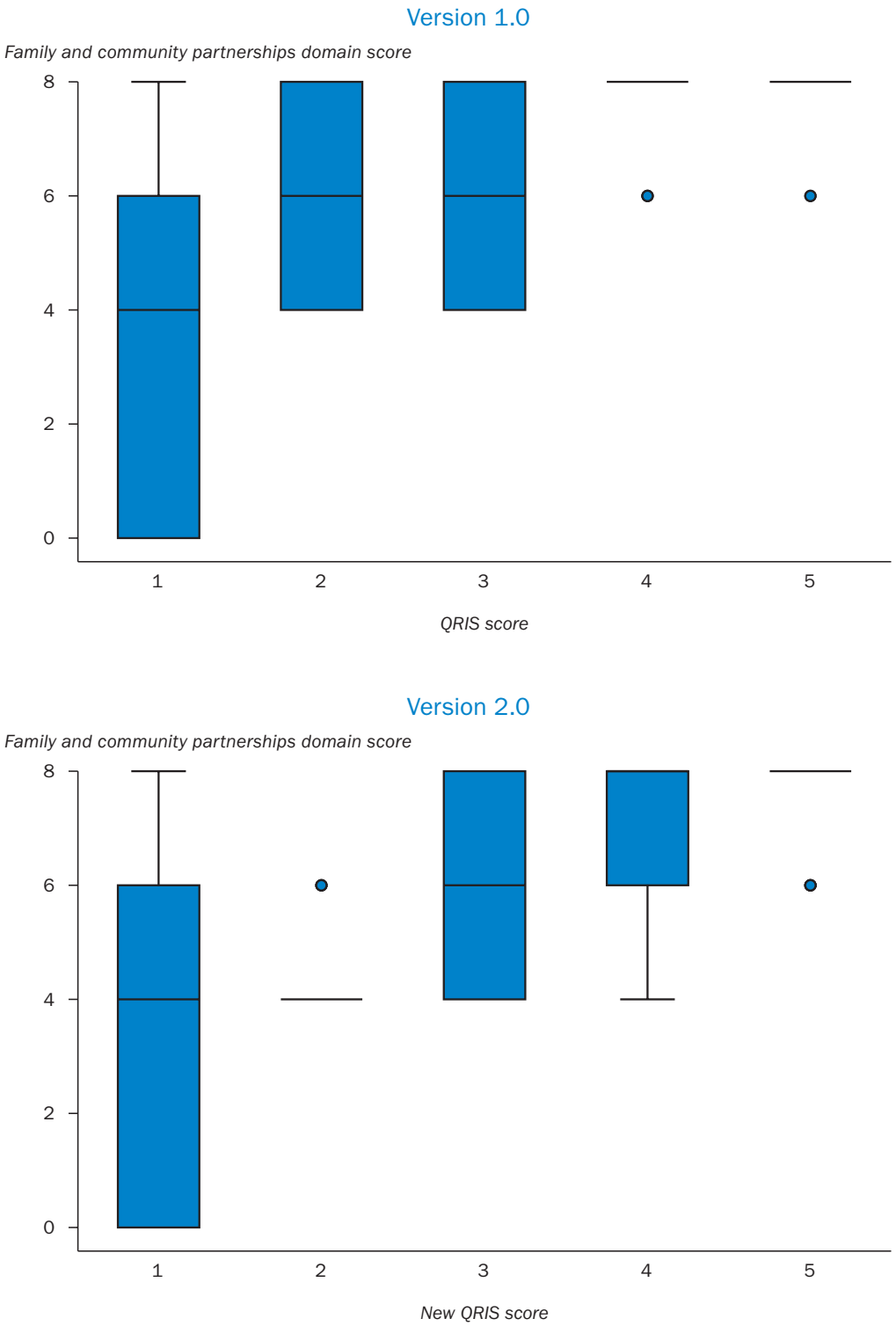
QRIS is quality rating and improvement system.

n = 1,413.

Note: The maximum value on the environment domain score is 8. Circles represent data points that are considered outliers. Data are as of January 16, 2013.

Source: Version 1.0, data provided by the Michigan Department of Education's Office of Great Start; version 2.0, authors' calculations using those data.

Figure F5. Box-and-whisker plots for the family and community partnerships domain of the Self-Assessment Survey, by QRIS score, under the version 1.0 and 2.0 approaches of Michigan's Great Start to Quality



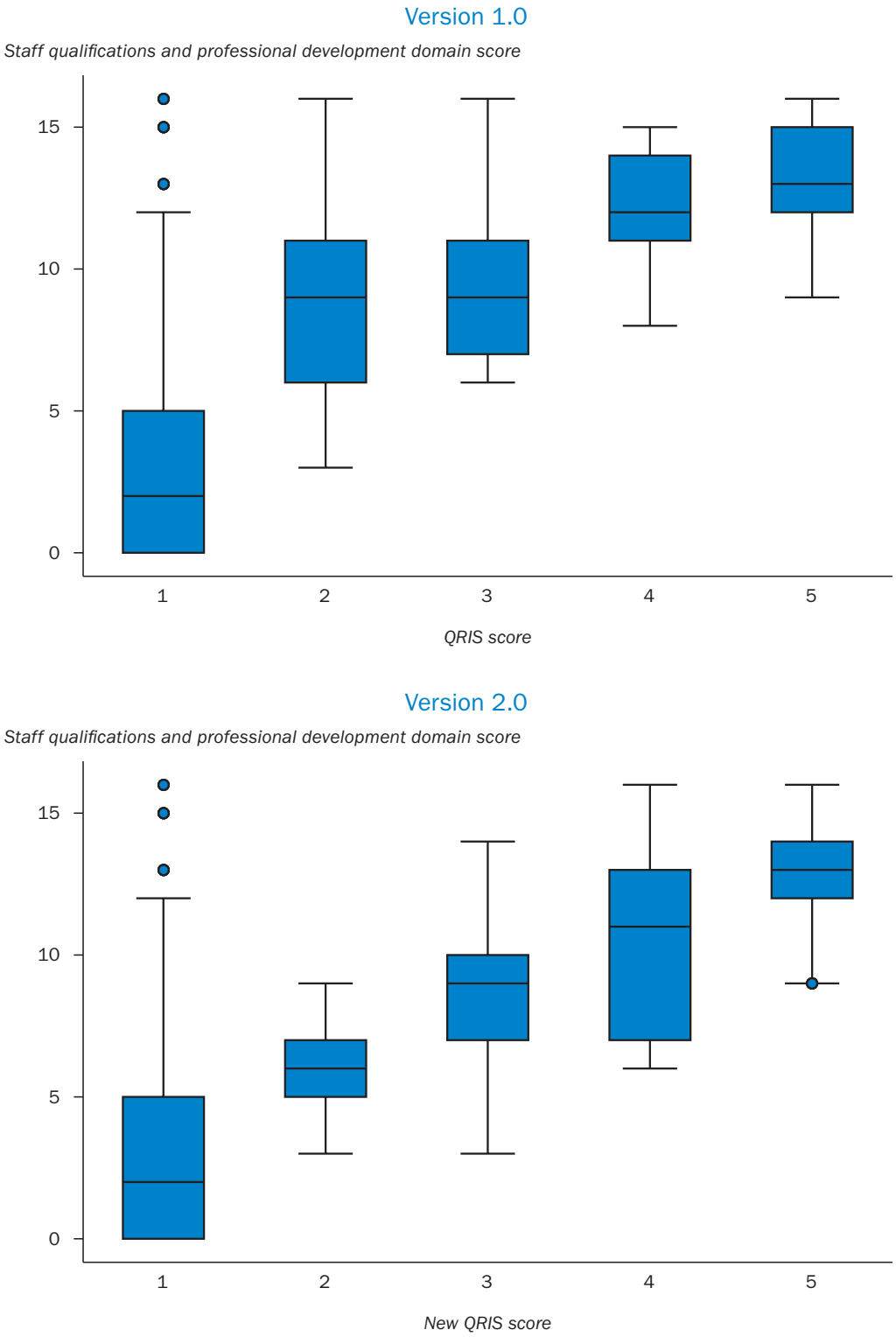
QRIS is quality rating and improvement system.

n = 1,413.

Note: The maximum value on the family and community partnerships domain score is 8. Circles represent data points that are considered outliers. Data are as of January 16, 2013.

Source: Version 1.0, data provided by the Michigan Department of Education's Office of Great Start; version 2.0, authors' calculations using those data.

Figure F6. Box-and-whisker plots for the staff qualifications and professional development domain of the Self-Assessment Survey, by QRIS score, under the version 1.0 and 2.0 approaches of Michigan's Great Start to Quality



QRIS is quality rating and improvement system.

n = 1,413.

Note: The maximum value on the staff qualifications domain score is 16. Circles represent data points that are considered outliers. Data are as of January 16, 2013.

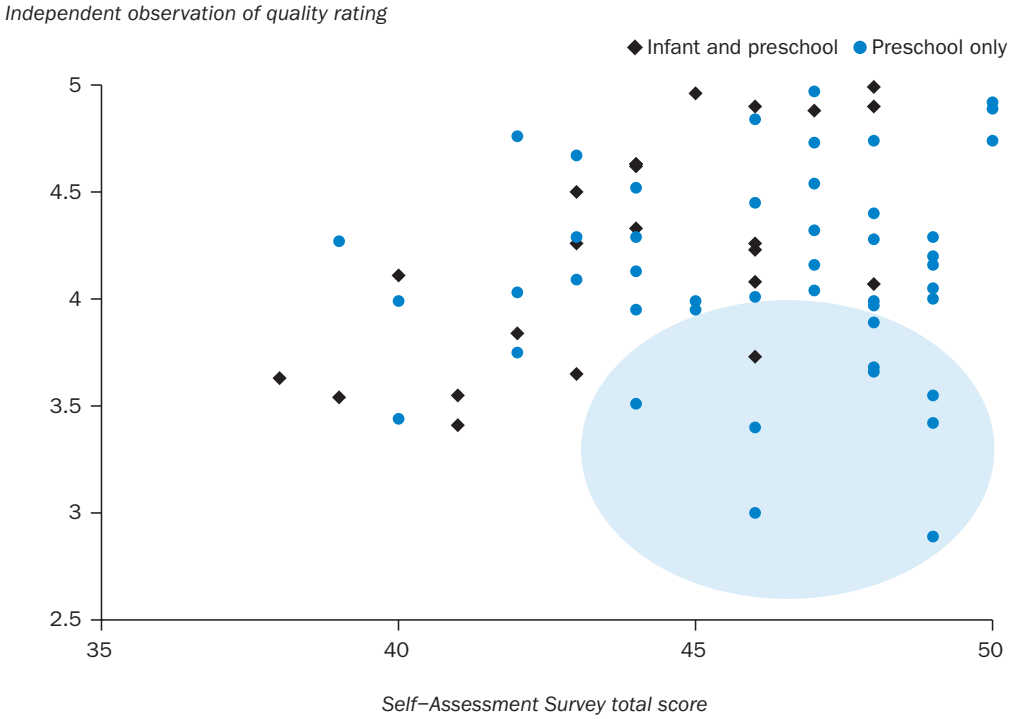
Source: Version 1.0, data provided by the Michigan Department of Education's Office of Great Start; version 2.0, authors' calculations using those data.

Appendix G. Associations between the Self-Assessment Survey and form A of the Program Quality Assessment

As described in the main report, the self-assessment ratings (from the Self-Assessment Survey) did not always result in the same rating as the independent observations of quality (from form A of the Program Quality Assessment) among the 72 programs with both ratings. The study team used two separate nonparametric measures of association, Spearman's ρ and Kendall's τ, to test the correlation between the two ratings. For these 72 programs, there was no significant association between programs' self-assessment ratings and independent observations of quality using either measure of association (Spearman's $\rho = .19$, $p = .11$, and Kendall's $\tau = .14$, $p = .10$). To better understand the lack of association, the study team examined a scatterplot of Self-Assessment Survey total scores and Program Quality Assessment scores. Although an overall positive linear relationship appears between the two sets of data, a pocket of programs with high Self-Assessment Survey total scores but lower ratings on the independent observations of quality exists (figure G1, blue shaded area).

Another concern with the measures of association between the scores on the self-assessment and independent observation instruments was the restricted range of scores in the sample of 72 programs. As described more fully in appendix F, the range of scores on

Figure G1. A scatterplot of Self-Assessment Survey total scores and ratings on the independent observation of quality shows a pocket of programs that have high Self-Assessment Survey total scores but lower ratings on the independent observation of quality



n = 72.

Note: Data are as of January 16, 2013.

Source: Data provided by the Michigan Department of Education's Office of Great Start.

both instruments was restricted to the higher end of the distribution, because Michigan administers form A of the Program Quality Assessment only to programs that self-rate at level 4 or 5. Coefficients for the measures of association should thus be interpreted with caution because they describe the association between scores on the self-assessment and independent observations instruments in a narrowly focused sample of only programs of high quality as rated on the Self-Assessment Survey. The coefficients cannot be generalized to all programs that participate in Great Start to Quality or to all programs in Michigan.

The restricted range was even more pronounced on the domains related to curriculum and environment. For example, programs that self-rated at level 4 or 5 had only three scores in the curriculum domain (8, 10, or 12) and two scores in the environment domain (6 or 8). For this reason, measures of association between the subscales of the Self-Assessment Survey and form A of the Program Quality Assessment were not conducted. Also, although the domain names are similar, the domains do not necessarily measure the same constructs (see tables C5 and C6 in appendix C).

Notes

1. Missouri requires legislative action to implement a QRIS.
2. These numbers were accurate as of 2010, when the most recent compendium of QRISs was released (Tout et al., 2010). Although these data are now five years old, no other comprehensive document describes the domains measured in all 49 QRISs.
3. Great Start to Quality is overseen by the Office of Early Learning at the Michigan Department of Education and the Early Childhood Investment Corporation. Rather than repeat these organizations, this report uses "Michigan" to refer to the key early childhood education stakeholders in the state.
4. To obtain licensing, early childhood education programs must demonstrate compliance with a set of requirements that ensure basic health, safety, and provider qualifications. For Michigan's licensing for child care centers, see Michigan Department of Human Services (2014); for Michigan's licensing for family and group child care homes, see Michigan Department of Human Services (2009).
5. For about 20 percent of programs in the study sample, the state verified the self-assessment ratings by reviewing program documentation and reassessing the preliminary rating level.
6. The Self-Assessment Survey instrument is not publicly available, so a table of items is not included in appendix C.

References

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology, 10*(4), 541–552.

Burchinal, M. R., Cryer, D., Clifford, R. M., & Howes, C. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science,* 6(1), 2–11.

Caronongan, P., Kirby, G., Malone, L., & Boller, K. (2011). *Defining and measuring quality: An in-depth study of five child care Quality Rating and Improvement Systems* (OPRE Report 2011–29). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. Retrieved July 1, 2013, from http://QRISnetwork.org/sites/all/files/resources/gscobb/2011–09–28%2014:01/Report.pdf

Christenson, S. L., & Reschly, A. L. (Eds.). (2010). *Handbook of school–family partnerships*. New York, NY: Routledge.

Fu, V. 1998. Estimating generalized ordered logit models. *Stata Technical Bulletin, 8*(44), 27–30. Retrieved July 1, 2013, from http://www.stata-journal.com/sjpdf.html?articlenum=gn0032

Harms, T., & Clifford, R. (1980). *Early childhood environment rating scale*. New York, NY: Teachers College Press.

High/Scope Educational Research Foundation. (2003). *Preschool program quality assessment, (PQA) administration manual* (2nd ed.). Ypsilanti, MI: High/Scope Press.

Lambert, R., Abbott-Shim, M., & Sibley, A. (2005). Evaluating the quality of early childhood educational settings. In B. Spodek & O. Saracho (Eds.), *Handbook of research on the education of young children* (2nd ed., pp. 457–475). Mahwah, NJ: Erlbaum. http://eric.ed.gov/?id=ED493785

Long, J. S., & Freese, J. (2006). *Regression models for categorical dependent variables using Stata* (2nd ed.) College Station, TX: Stata Press.

Mashburn, A. J. (2008). Quality of social and physical environments in preschools and children's development of academic, language, and literacy skills. *Applied Developmental Science, 12*(3), 113–127. http://eric.ed.gov/?id=EJ887977

Michigan Department of Human Services. (2009). *Licensing rules for family and group child care homes*. Lansing, MI: Author. Retrieved July 1, 2013, from http://www.michigan.gov/documents/dhs/BCAL-PUB-0724_212611_7.pdf

Michigan Department of Human Services. (2014). *Licensing rules for child care centers*. Retrieved July 1, 2013, from http://www.michigan.gov/documents/dhs/Child_Care_Center_Rules_419095_7.pdf

National Institute of Child Health and Human Development 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*(3), 199–206. http://eric.ed.gov/?id=EJ654421

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., et al. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science, 9*(3), 144–159. Retrieved July 1, 2013, from http://www.leg.state.vt.us/PreKEducationStudyCommittee/Documents/impact%20of%20teacher%20and%20classroom%20quality.pdf

QRIS National Learning Network. (2007). *A great start for kids: Recommendations for a Michigan child care quality rating and improvement system*. Retrieved July 1, 2013, from http://greatstartforkids.org/sites/default/files/file/QRS_FinalB_2007.pdf

QRIS National Learning Network. (2014). *QRIS state contacts & map*. Retrieved November 28, 2014, from http://qrisnetwork.org/sites/all/files/contacts/QRIS%20State%20Contacts,%20QRIS%20National%20Learning%20Network,%20www.qrisnetwork.org%20[Revised%20November%202014]_0.xlsx

Sabol, T. S., Soliday Hong, S. L., Pianta, R. C., & Burchinal, M. R. (2013). Can rating pre-k programs predict children's learning? *Science, 321*(4), 845–846. Retrieved August 30, 2013, from http://www.sciencemag.org/content/341/6148/845.short

Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2010). *Compendium of quality rating systems and evaluations*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. Retrieved July 1, 2013, from http://www.childcareresearch.org/childcare/resources/18554/pdf

Tout, K., Chien, N., Rothenberg, L., & Li, W. (2014). *Implications of QRIS design for the distribution of program ratings and linkages between ratings and observed quality* (OPRE Research Brief #2014–33). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. Retrieved November 7, 2014, from http://www.acf.hhs.gov/sites/default/files/opre/ecls_brief1.pdf

U.S. Department of Education. (2011). *Race to the Top-Early Learning Challenge: Applicant information*. Washington, DC: Author. Retrieved July 1, 2013, from http://www2.ed.gov/programs/racetothetop-earlylearningchallenge/applicant.html

U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. (2010). *Ohio Step Up to Quality: QRS profile*. Washington, DC: Authors. Retrieved July 1, 2013, from http://www.acf.hhs.gov/programs/opre/cc/childcare_quality/ohio/ohio.pdf

Washington State Department of Early Learning. (2013). *Early Achievers, Washington's quality rating and improvement system standards: A framework to support positive child*

outcomes. Olympia, WA: Author. Retrieved July 1, 2013, from http://www.del.wa.gov/publications/elac-qris/docs/EA_Quality_Standards.pdf

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal,* 6(1), 58–82. Retrieved July 1, 2013, from http://www.stata-journal.com/sjpdf.html?articlenum=st0097

Zellman, G. L., & Fiene, R. (2012). *Validation of quality rating and improvement systems for early care and education and school-age care* (Research-to-Policy, Research-to-Practice Brief OPRE 2012–29). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. http://eric.ed.gov/?id=ED534457

Zellman, G. L., & Perlman, M. (2008). *Child-care quality rating and improvement systems in five pioneer states: Implementation issues and lessons learned.* Santa Monica, CA: RAND Corporation. Retrieved July 1, 2013, from http://www.rand.org/content/dam/rand/pubs/monographs/2008/RAND_MG795.pdf

The Regional Educational Laboratory Program produces 7 types of reports

Making Connections
Studies of correlational relationships

Making an Impact
Studies of cause and effect

What's Happening
Descriptions of policies, programs, implementation status, or data trends

What's Known
Summaries of previous research

Stated Briefly
Summaries of research findings for specific audiences

Applied Research Methods
Research methods for educational settings

Tools
Help for planning, gathering, analyzing, or reporting data or research