Great Start to Quality Program Validation Study Final Report

A project funded by the Michigan Department of Education with funding from a US Department of Education Race to the Top — Early Learning Challenge Grant









Acknowledgments

HighScope Educational Research Foundation and the American Institutes for Research would like to extend their sincere appreciation to the many people who provided critical feedback, perspectives, and helpful suggestions to this study. In particular, we would like to thank the Michigan Department of Education, members of the expert panel, and the directors, owners, teachers, and support staff from the early care and education programs around the state who contributed the information required for this research. We would also like to thank Joy Milano and all of the data collectors who were vital to ensuring the study was carried out with the highest degree of quality and success.

Project Staff

HighScope Educational Research Foundation	Americ
Iheoma U. Iruka, PI	Ann-M
Jill Claxton, Project Manager	Natalie
Fernando Andrade, Methodologist	Jingtor
Carol Markley, Project Coordinator	Emma
Beth Hardin, Project Coordinator	Shahee
Reino Bruner, Data Scientist	

American Institutes for Research Ann-Marie Faria, Co-PI Natalie Tucker-Bradway, Analyst Jingtong Pan, Methodologist Emma Nyhof, Research Assistant Shaheen Khan, Data Scientist

Report Information

This study was funded by the Michigan Department of Education with funding from a US Department of Education Race to the Top – Early Learning Challenge Grant. The study was conducted by HighScope Educational Research Foundation in collaboration with the American Institutes for Research.

About HighScope Educational Research Foundation

The HighScope Educational Research Foundation has a 40-year history of conducting research while ensuring professionalism in the field of early childhood. HighScope is an independent, nonprofit research and practice organization committed to supporting children and educators by providing high-quality early childhood education. HighScope is well-known for conducting the landmark Perry Preschool Study that first established the lasting human and financial value of high-quality early childhood education.

About American Institutes for Research

Founded in 1946, AIR is one of the largest not-for-profit behavioral and social science research and evaluation organizations in the world. AIR is committed to empowering communities and institutions with innovative solutions to the most critical challenges in education, health, workforce productivity, and international development.

Suggested citation: Iruka, I. U., Faria, A.-M., Claxton, J., Tucker-Bradway, N., Andrade, F., Pan, J., Markley, C., Hardin, B., Bruner, R., Nyhof, E., & Khan, S. (2018). *Great Start to Quality Validation Study Final Report.* Ypsilanti, MI: HighScope Educational Research Foundation.

TABLE OF CONTENTS

I.	EXECUTIVE SUMMARY5
н.	STUDY PURPOSE AND RESEARCH QUESTIONS7
	a. Overview7
	b. GSQ and Rating Process7
	c. Synthesis of QRIS Validation Study9
	d. Overview of Report10
III.	METHODOLOGY 11
	a. Sampling Frame11
	b. Program-Level Recruitment12
	c. Extant Data Sample15
	d. Sample15
	e. Measures16
	f. Data Collector Training and Quality Control18
IV.	RESULTS
	1.1 Are the QRIS quality components and standards the "right" ones? 21
	1.2 Is the process of documenting and verifying each indicator yielding
	reliable and accurate results?25
	1.3 What is the relationship among the components and are they functioning
	as expected? That is, how do the SAS and the PQA correlate to each
	other and to the overall rating of quality?
	1.4 What are the most appropriate ways to combine measures of quality
	standards into summary ratings when examining the cut scores and
	combining rules?
	1.5 What is the variation and pattern of program-level ratings within and
	across program types that ensures the ratings are functioning as
	intended?50
	2.1 Are there differences in GSQ ratings based on other measures of quality
	using the Environment Rating Scales (ERS) and Classroom Assessment
	Scoring System (CLASS) across program types?
	2.2 How well do the ratings from the alternate cut points and rules
	distinguish different levels of quality?59
v.	Discussion
	a. Limitations
VI.	Recommendations65
VII.	References
VIII.	Appendix

TABLE OF TABLES

Table 2-1. Non-Alternate Path Providers' Participation Counts: Licensing Type by Published Rating	9
Table 2-2. GSQ Validation Questions and Data Sources	10
Table 3-1. Number of Programs That Passively Declined to Participate	14
Table 3-2. Number of Programs That Actively Declined to Participate	14
Table 3-3. Reasons Programs Actively Declined to Participate	14
Table 3-4. Number of Programs Participating by Program Type and Star Rating	16
Table 3-5. Number of Classrooms Participating by Program Type and Star Rating	16
Table 4-1. Distribution of Relevance of Quality Indicators by Area and Across Raters	22
Table 4-2. Top 4 Indicators of SAS That Achieved Highest Rating for Child Outcomes and Program	
Quality	23
Table 4-3. Program Type and Star Rating for Reliability Sub-Study	25
Table 4-4. Summary Statistics for SAS Validation Ratings by License Type per Area	26
Table 4-5. Correlation Matrix Between HighScope Rating and MDE Rating of the Same Programs	27
Table 4-6. Comparison of Self-Rating and Independent Observation of Quality	28
Table 4-7. Correlations Among the Final SAS Rating, Observed PQA Score, and Published GSQ	
Rating	29
Table 4-8. Correlation of SAS Sub-Domains Scores With Each Other and Self-Assessment Ratings	31
Table 4-9. Person Separation Reliability, Internal Consistency, Range of Item Fit, and Point-	
Correlation Range of SAS Program Quality Domains	34
Table 4-10. Indicators Included in the Streamlined Approach Versus the Original SAS	40
Table 4-11. Three Possible Cut Scores Using the Streamlined Approach for Centers and Homes	44
Table 4-12. Correlation ERS and Ratings	56
Table 4-13. Summary of ANOVA Analyses Examining Differences in ERS by Program Type and Star	
Rating	57
Table 4-14. CLASS and ERS Scores by Level Rating in the New Streamlined Approach	60
Table 4-15. CLASS and ERS Scores by Level Rating in the New Few and Mighty Approach	62
Table A1. Domains, Items, and Response Rates for Center-Based Classrooms and Home-Based	
Classrooms	71
Table A2. Determining the Goodness of Fit for Different Factor Structures (Combined)	73
Table A3. Factor Loadings for the Five-Factor Model (Combined)	74
Table A4. Determining the Goodness of Fit for Different Factor Structures (Center-Based)	76
Table A5. Factor Loadings for the Five-Factor Model (Center-Based)	77
Table A6. Determining the Goodness of Fit for Different Factor Structures (Home-Based)	78
Table A7. Factor Loadings for the Two-Factor Model (Home-Based)	79
Table A8. Construct Reliability and Properties	81
Table A9. SAS Items Listed From Easiest to Hardest (Combined Sample)	82
Table A10. Easy Items That Do Not Differentiate Quality in Any Programs by Sample Type	83
Table B1. Comparison of Participation in GSQ by Key Program Characteristic	84
Table B2. Comparison of All Michigan Early Childhood Programs and Programs Participating in	
GSQ, by Region	85
Table B3. Comparison of Programs Participating in GSQ, by Program Type and Other Key	
Characteristics	85
Table B4. Sample Comparison of SAS and Published GSQ Ratings	86
Table B5. Final GSQ Ratings by Program Characteristics	86

TABLE OF FIGURES

Figure 2-1. GSQ Star Classification Based on Total Points Earned	8
Figure 3-1. Selected Sample by Program Type and Star Rating	12
Figure 4-1. Single-Factor Rasch Analysis Relative to Self-Assessment of Program Quality	36
Figure 4-2. Domains and Number of Indicators in the Original and Streamlined SAS	38
Figure 4-3. Domains and Number of Indicators in the Original and Streamlined SAS	39
Figure 4-4. Distribution of Total Streamlined SAS Scores by Provider Type	43
Figure 4-5. Center Ratings Using the Higher Cut Scores and Home Ratings Using the Lower Cut	
Scores for the Streamlined Approach	45
Figure 4-6 Number of Programs Rated as Low, Moderate, and High Quality in the Original SAS	
Versus the Streamlined Approach, for Centers and Homes	46
Figure 4-7. The Few and Mighty Indicators	47
Figure 4-8. The Percentages of Center- and Home-Based Programs That Met and Did Not Meet All	
Few and Mighty Indicators for the "High Quality" Rating	48
Figure 4-9. Final Published GSQ Ratings	51
Figure 4-10. Final Published GSQ Ratings by Program Type	52
Figure 4-11. SAS Ratings by Program Type	53
Figure 4-12. Environment Ratings Scale Ratings Across Type of Tool	54
Figure 4-13. Average CLASS Domain Scores	55
Figure 4-14. Environment Ratings Scale Ratings by Star Rating	57
Figure 4-15. Average CLASS Domain Scores by Star Rating	58
Figure A1. Eigenvalues of the Exploratory Factor Analysis, SAS (combined)	72
Figure A2. Eigenvalues of the Exploratory Factor Analysis, SAS (Center-Based)	75
Figure A3. Eigenvalues of the Exploratory Factor Analysis, SAS (Home-Based)	78

I. EXECUTIVE SUMMARY

This report presents the results of a validation study of Michigan's Quality Rating Improvement System (QRIS), Great Start to Quality (GSQ), using a sample of participating center- and homebased early care and education (ECE) providers. HighScope Educational Research Foundation (HighScope), in collaboration with the American Institutes for Research (AIR), conducted this study in 2018. The project examines the following research questions: (1) if and how GSQ measures "quality" for licensed centers and home-based providers; and 2) the difference between ratings in terms of quality. Specifically, this study examines whether the self-assessing criteria reveal meaningful distinctions in quality between programs and if differences in GSQ ratings are supported by two other measures of quality, the Environment Rating Scales (ERS) and Classroom Assessment Scoring System (CLASS).

GSQ is a 5-level system that measures the quality of early childhood programs and providers in Michigan by using more than 40 program quality indicators aligned with Michigan's Early Childhood Standards of Quality for Infant and Toddler Programs and Early Childhood Standards of Quality for Prekindergarten. The program quality indicators are standards used to measure the quality of a program that fall into five categories: Staff Qualifications and Professional Development; Family and Community Partnerships; Administration and Management; Environment; and Curriculum and Instruction. Programs can earn a maximum of 50 points on the GSQ rating scale ranging from 1- to 5-stars.

Based on current research linking high-quality ECE experiences to children's school readiness and later success, QRIS have been developed to enhance and ensure that ECE programs are high quality, especially for children most at-risk for school failure (e.g., low income, ethnic minority, dual language learners). As QRIS have multiple functions — including providing a systematic way to monitor and improve quality and providing clear information to families about quality — there is a need to ensure that the system is, foremost, valid and reliable. Sampling for this evaluation was based on all eligible licensed programs in the state participating in GSQ with a star rating of 1 to 5. To truly understand the validity of GSQ as intended, we focus on programs that don't have an alternate pathway to reach the highest levels. Great Start Readiness, Head Start, and NAEYC-accredited programs have an alternate pathway to reach the highest levels in GSQ; thus, they were not eligible to be part of this study. The validation study team recruited a sample of 72 child care centers, 52 group family child care homes, and 58 family child care homes, totaling 182 programs. In these 182 programs, observation data were collected across 306 classrooms.

To empirically validate the GSQ ratings, valid and reliable global observation measures of quality were used – Environment Rating Scales (ERS) for centers, family child care homes, and group homes, and Classroom Assessment Scoring System (CLASS) for pre-K classrooms at centers. Results indicated that programs at higher star ratings (i.e., 4- and 5-star) scored better than programs at lower star ratings on these global measures of quality. While there were no differences by program type in total ERS scores, there were areas in the ERS where family child care homes scored higher than centers (i.e., Language and Literacy and Learning Activities); and areas where center programs scored higher (Personal Care Routines). There was a trend that programs with higher star ratings (i.e., 4- and 5-star) scored better than 3-star programs in the

CLASS domains of Emotional Support and Classroom Organization; there were no significant differences by rating on Instructional Support, which is the domain found to be related to children's outcomes.

Extant data analyses of all GSQ programs found that the GSQ's Self-Assessment Survey (SAS) was not measuring the underlying aspects of quality as intended. The SAS protocols include many "easy" items for which nearly all programs receive points, reducing the psychometric validity and reliability of the measure. Based on this finding, two alternate structures for scoring the SAS were proposed, a streamlined and a "Few and Mighty" approach. The indicators included in these alternate structures were based on the extant analyses as well as expert panel and director survey feedback. The streamlined approach consistently identified high-quality programs from moderate- or low-quality programs and works best for home-based providers; however, more centers are still identified as high quality. The ratings from the Few and Mighty approach do well to distinguish two distinct levels of quality for home-based programs, but do not for center-based programs. Home-based programs rated as High Quality on the Few and Mighty approach have statistically significant and higher ERS scores than those rated as Improving Quality. Centers rated as High Quality had higher CLASS and ERS scores than Moderate and Low Quality programs, but they were not statistically significant. This may be explained in part due to the small sample sizes for center-based programs. The streamlined and Few and Mighty approaches seem to be picking up on meaningful differences in quality for home-based programs as measured by the ERS, but may not be as precise for centers.

The GSQ system is, overall, a reliable system where generated ratings are aligned with the MDE standards for quality; however, there is a concern that some components of the system may not be psychometrically sound, which requires deeper analysis as to what aspects are most critical to improve program quality and support children's development and learning. This study should be helpful to Michigan as they seek ways to improve quality in center- and home-based programs across the state; however, caution should be taken in generalizing the findings in this study due to the small sample size and the non-inclusion of child outcomes.

II. STUDY PURPOSE AND RESEARCH QUESTIONS

a. Overview

This report presents the results of a validation study of Michigan's Quality Rating Improvement System (QRIS), Great Start to Quality (GSQ), using a sample of participating center- and homebased early care and education (ECE) providers. HighScope Educational Research Foundation (HighScope), in collaboration with the American Institutes for Research (AIR), conducted this study in 2018. The project examines the following research questions: (1) if and how GSQ measures "quality" for licensed centers and home-based providers; and 2) the difference between ratings in terms of quality. Specifically, this study examines whether the self-assessing criteria reveal meaningful distinctions in quality between programs and if differences in GSQ ratings are supported by two other measures of quality, the Environment Rating Scales (ERS) and Classroom Assessment Scoring System (CLASS).

b. GSQ and Rating Process

Michigan developed its QRIS, Great Start to Quality (GSQ), in the early 2000s and rolled out the system statewide in 2012, with more than 2,000 programs participating by the end of that year. GSQ uses a hybrid rating calculation approach that draws on a self-assessment of structural quality and independent observations of quality (for programs that want to achieve 4- or 5-star rating), both of which are based on the Program Quality Assessment (PQA). In January 2013, Michigan decided to revise its approach to calculating ratings to apply different cut scores on the self-assessment, change the requirements for earning points on the staff qualifications subdomain, and revise the standards for home-based programs. The ratings range from level 1 to level 5, where level 1 indicates the lowest quality and level 5 indicates the highest quality. GSQ includes a variety of types of licensed early childhood education programs (center, group home, family home) that serve children from birth to age 5. (These programs also serve schoolage children, but they are not a focus of this study.) GSQ is a voluntary system that publishes ratings so parents can easily identify the best early care and education (ECE) options for their children and providers have clear quality standards. GSQ includes incentives for programs to participate, such as public ratings for high-quality programs, tiered reimbursement based on GSQ rating, and eligibility for special grant funding opportunities (dependent on availability of funding).

GSQ measures the quality of early childhood programs and providers in Michigan by using more than 40 program quality indicators aligned with Michigan's Early Childhood Standards of Quality for Infant and Toddler Programs and Early Childhood Standards of Quality for Prekindergarten. The program quality indicators are standards used to measure the quality of a program in a specific area. Each program quality indicator falls into one of five categories:

- Staff Qualifications and Professional Development the training and any education or degrees completed by providers and their staff
- Family and Community Partnerships how programs offer support to and interact with the families and communities they serve

- Administration and Management how policy guides the way a program is run
- Environment the safety of a program and how the program supports the health and wellness of the children
- Curriculum and Instruction the tools a program uses to teach, accommodate needs, and track development of the children in their care

The GSQ star classification based on total points earned is shown in Figure 2-1.

	Statewide
★	Provider meets licensing requirements and is participating in GSQ
**	Minimum of 16 Points in 2 of 5 Categories
***	Minimum of 26 Points in 3 of 5 categories
***	Minimum of 38 Points in 4 of 5 categories
	PQA score higher than 3.5
****	Minimum of 42 Points in 5 of 5 categories
	PQA score higher than 4.5

Figure 2-1. GSQ Star Classification Based on Total Points Earned

Table 2-1 shows the distribution of centers, group and family child care homes, and tribal centers by star rating (as of June 14, 2018) that are not Great Start to Readiness, Head Start, and NAEYC-accredited programs (also known as non-alternate path providers).

Table 2-1. Non-Alternate Path Providers' Participation Counts: Licensing Type by PublishedRating

License Type	Statewide Provider Pool	1 STAR	2 STAR	3 STAR	4 STAR	5 STAR	Published 1 to 5 Stars
Licensed Center	1,950	2	33	589	108	26	758
Licensed Group Home	1,587	35	121	485	22	28	691
Registered Family Home	2,900	94	299	562	18	15	988
Tribal Center	3	0	0	2	0	0	2
Total Programs	6,440	131	453	1,638	148	69	2,439

c. Synthesis of QRIS Validation Study

Three key promises of a QRIS are 1) to ensure that parents and other stakeholders can select the highest quality child care programs for children based on meaningful quality ratings; 2) to promote quality improvement in all child care settings through the provision of important benchmarks of quality measured periodically to examine change over time; and 3) to provide quality that can support children's optimal development (Zellman & Fiene, 2012). In order to make good on these promises, a QRIS must be validated. Validation is a multi-faceted process that involves demonstrating a high degree of correlation between ratings and important indicators of program quality (construct validity), a high degree of correlation between ratings and important measures of quality not included as indicators in the rating system (convergent validity), a high degree of correlation between quality ratings and desired child outcomes (predictive validity), and a significant capacity of the ratings system to distinguish high- and lowquality sites (discriminant validity). In addition, the validation process assumes reliability in quality measures, an assumption that requires indicators and quality ratings to not be biased or flawed in their reflection of quality status. Given these definitions, we conducted a multipronged validation of GSQ that is framed by Zellman and Fiene's (2012) comprehensive approach to validating a quality rating and improvement system and includes

- Examining the validity of the underlying concepts of the system
- Evaluating the psychometric properties of the elements of the system (e.g., correlation among standards and indicators)
- Assessing the outputs of the rating system (e.g., relation with independent measures)

While we seek to meet the criteria for validation of GSQ as specified by Zellman and Fiene (2012), we addressed aspects focused on examining the validity of the underlying concepts of GSQ (namely, the 5 categories and indicators of the Self-Assessment Survey [SAS]) and assessed the outputs of the rating system in relation to other measures. Examination of the association between ratings and child outcomes — another step in the validation process — was not included in this evaluation due to timing and budget constraints. Our evaluation took into

account the mixed findings about QRIS, including few programs at the highest levels of the system and minimal focus on home-based providers.

d. Overview of Report

The Methodology, including sampling strategy, data collection methods, and data collector training and quality control, is presented in the next section, followed by a Results section, and finally Discussion and Recommendations. Table 2-2 displays a summary of the key validation study questions and the sources of data used to answer each question. In Section 5, we summarize the results of the study and implications for the GSQ QRIS.

Table 2-2. GSQ Validation Questions and Data Sources

Validation Questions	Data Source
1. If and how does GSQ actually measure "quality" in licensed centers and for home-based providers?	
1.1 Are the QRIS quality components and standards the "right" ones?	Expert survey
1.2 Is the process of documenting and verifying each indicator yielding reliable and accurate results?	Administrative data
1.3 What are the relationships among the components and are they functioning as expected?	Administrative data
1.4 What are the the most appropriate ways to combine	Administrative data
measures of quality standards into summary ratings when	Expert survey
examining the cut scores and combining rules?	Director survey
1.5 What is the variation and pattern of program-level ratings within and across program types that ensures the ratings are functioning as intended?	Administrative data
2. What is the difference between ratings in terms of quality? Specifically, do the self-assessing criteria produce meaningful distinctions in quality?	
2.1 Are there differences in GSQ ratings based on other	Administrative data
measures of quality using the Environment Rating Scales	Family child care quality
(ERS) and Classroom Assessment Scoring System (CLASS)	observations
	Classroom quality observations
2.2 How well do the ratings from the alternate cut points	Administrative data
and rules distinguish different levels of quality?	Family child care quality
	observations
	Classroom quality observations

III. METHODOLOGY

This section describes the sampling and recruitment strategies for child care centers, group centers, and family child care homes. In addition, the program response rate is reported.

a. Sampling Frame

We based our sampling approach on eligible programs' participation in GSQ as of June 14, 2018, as indicated in Table 2-1 (i.e., non-alternate pathway centers, group child care homes, and family child care homes). Programs were eligible to participate in the validation study if they were participating in GSQ and had a published star rating as of June 2018. (See Table 2-1 for information on the types of programs.) We used a stratified random sampling to select 450 programs that were representative of Michigan early childhood education's diverse regions/locales, quality ratings, care by provider type, and for-profit/nonprofit status. The most important strata were program type (center, group, and family child care homes) and GSQ rating level, for which we tried to sample an equal number of programs within each star level, where possible. In instances where there were not enough of a specific type to sample from (such as center-based programs at the 1-star level), we selected all programs.

We used a **stratified probability sampling** that prioritizes the stratification of program type and star rating at each level (see Figure 3-1). The 15 strata were:

- 1. Center-based Programs with 1-star rating
- 2. Center-based Programs with 2-star rating
- 3. Center-based Programs with 3-star rating
- 4. Center-based Programs with 4-star rating
- 5. Center-based Programs with 5-star rating
- 6. Group Child Care Homes with 1-star rating
- 7. Group Child Care Homes with 2-star rating
- 8. Group Child Care Homes with 3-star rating
- 9. Group Child Care Homes with 4-star rating
- 10. Group Child Care Homes with 5-star rating
- 11. Family Child Care Homes with 1-star rating
- 12. Family Child Care Homes with 2-star rating
- 13. Family Child Care Homes with 3-star rating
- 14. Family Child Care Homes with 4-star rating
- 15. Family Child Care Homes with 5-star rating

The goal was to include a minimum of 30 programs in each stratum (star rating and program type), which ensures that selection into the study is entirely by chance. In Figure 3-1, we show our expected sample by program type and rating. We highlight strata where there are fewer than 30 programs in red text and those with more than 30 in blue.





b. Program-Level Recruitment

One of the initial recruitment steps was an email from MDE, sent on June 15, 2018, to encourage programs to participate in this validation study. The email was distributed through MDE listservs and partner listservs as indication that MDE authorized this study and provided information about how the study offers opportunities for strengthening the GSQ system. Provider engagement was stressed as the way to most accurately reflect program quality. Four additional email messages (June 30, 2018–October 23, 2018) were sent from MDE throughout the study to encourage program participation.

Upon Institutional Review Board (IRB) approval, recruitment of programs commenced on August 17, 2018 with an email informing programs they had been selected for study participation. The email included information about the purpose of the study, information about incentives for participation, and a link to a five-minute animated video that further explained the study. The video link was also posted on HighScope and MDE websites and shared on social media and communication platforms of relevant partners, such as the Early Childhood Investment Corporation (ECIC) and GSQ Resource Centers.

The direct communication Resource Centers have with childcare programs was leveraged. More specifically, Resource Center Directors (RCDs) were emailed information about the study to share with quality improvement consultants (QICs) and quality improvement specialists (QISs) asking that they discuss this study with programs as they visited or at meetings that program directors and owners attended. On October 26, 2018, HighScope met with the 11 RCDs to further explain the study, present the current participation status, and distribute lists to each director of programs (within their respective regions) that had been selected for study inclusion. The RCDs were asked to specifically reach out to the programs in their area that had been selected and encourage them to participate in the study.

Incentives were used to encourage participation in this project and to thank participants for their time. A study liaison was established at each participating program. Study liaisons supported the collection of consents and helped with data collection activities on site. Each study liaison received a \$50.00 gift card. Directors and owners who completed a program

survey received a \$20.00 gift card and classroom teachers and family child care providers received a \$20.00 incentive for allowing observations to take place in their classroom/home.

The next steps included a call to all selected programs requesting their participation. If no one answered, a message was left about the study and how to contact a HighScope staff person to participate. In addition, programs received a follow-up email either with a request for classroom information (if the program had already agreed to participate) or contact information for HighScope staff to inquire about participation in the study (if a message had been left). For programs that did not respond to the initial email, first call, or follow-up email, a second round of calls/emails was completed. Finally, programs that had not responded to calls and emails were sent a postcard with study information, how they could participate, and a request to call a HighScope staff member to participate.

Once a program agreed to participate, a random selection of classrooms was drawn, up to 4 classrooms per program, and consent was sought from classroom teachers. If a program had fewer than 4 classrooms, all classrooms were selected for participation. This strategy was followed for the initial sample and the two subsequent replacement samples. For the third replacement sample, the initial email as well as a postcard inviting programs to participate were sent simultaneously. Then all programs were called and sent a follow-up email.

One of the stumbling blocks with the initial phases of recruitment of programs was that many programs (60+) replied with a "maybe" response. Some of the responses include: "The beginning of the year is a bad time to do observations"; and "Call me back in a month, once my classrooms are settled and the children and teachers know the routine." This response was heard equally often from centers and family child care homes. Directors and owners were very reluctant to have observers come out to their programs at the start of a new school year. Other problems plagued the recruitment process, including other projects and studies being conducted at the same time, with the same population of programs as this one; and child care licensing requirements including new background check procedures, among other things. The largest problem, however, was programs not responding at all to requests for participation. Table 3-1 shows the breakdown by program type and star level of programs that passively declined — that is, never responded to any calls, emails, or postcard mailings. Table 3-2 shows the number of programs by program type and star level that actively declined to participate in the study. Table 3-3 further explains the reasons given for actively declining to participate. As indicated, an additional 46 programs initially agreed to participate in the study, but they either did not respond to the data collector to set up observations after several attempts to do so, cancelled the agreed-upon observation times, or called the recruiter back to state they were no longer interested in participating in the study. Also a large number of programs (39) had closed their programs altogether or were temporarily closed to reconstruct the building. This led to a reduction in the numbers of programs where data collection was anticipated.

License Type	1 STAR	2 STAR	3 STAR	4 STAR	5 STAR	Total	
Licensed Center	1	16	102	49	8	176	
Licensed Group Home	21	76	113	8	10	228	
Registered Family Home	41	66	118	7	2	234	
Total Programs	63	158	333	64	20	638	
Note. Data as of December 31, 2018							

Table 3-1. Number of Programs That Passively Declined to Participate

Table 3-2. Number of Programs That Actively Declined to Participate

License Type	1 STAR	2 STAR	3 STAR	4 STAR	5 STAR	Total	
Licensed Center	1	13	30	36	11	91	
Licensed Group Home	12	33	65	8	13	131	
Registered Family Home	41	30	51	6	7	135	
Total Programs	54	76	146	50	31	357	
Note. Data as of December 31, 2018							

Table 3-3. Reasons Programs Actively Declined to Participate

Reasons for Declining to Participate	Number of Programs
Declined/Not interested/Does not want an observation	176
Initially agreed but did not follow through with observation/Changed mind	46
Business closed/Moving/Under construction	39
Provider health issues/Personal issues	20
Program issues/New staff	18
Other program demands/Recent PQA/GSQ rating in process	17
Dissatisfaction with GSQ/MDE	5
Not interested in HighScope	3
GSRP program (excluded for study)	2

c. Extant Data Sample

Two samples were created using data provided by Michigan Department of Education on June 14, 2018 to complete all descriptive and psychometric analyses. The initial datafile included information on 9,243 licensed center- and home-based providers across the state of Michigan.

The first sample, the descriptive sample, included all programs participating in the GSQ to answer RQ 1.3 and 1.5. This sample included 3,796 programs that had a published GSQ rating and an additional 25 programs that had a completed SAS and were in the process of receiving a final GSQ rating.

The second sample, the psychometric and alternate score sample, included all SAS records with complete indicator level data. This included 4,397 completed SAS responses.¹ The sample for all analyses using indicators — psychometric analyses and analyses that created alternate scores — excluded programs with missing indicator data.²

d. Sample

An initial random sample of 520 programs was selected for inclusion in the GSQ Validation Study from the 2,439 programs participating in GSQ as of June 2018. The sample was stratified across license type and published star rating (as shown in Figure 3-1). An oversampling strategy was employed to ensure a minimum sample size of 450 programs would be reached.

Three replacement samples were drawn throughout the study period to take the place of programs that had declined (actively) to participate or had been contacted 5 or more times with no reply (passively declined). On October 10, 2018, 89 programs that had declined to participate were replaced, employing a 15% oversampling strategy. On October 23, 76 programs that had been contacted 5 or more times were replaced, employing a 60% oversampling strategy. The final replacement sample was drawn on November 9, 2018, with 205 programs (both active and passive declines) replaced, employing a 100% oversampling strategy. In total 1,177 programs were selected for inclusion in the study.

Table 3-4 reports the final sample of programs participating by program type and star rating. Table 3-5 reports the number of classrooms participating by program type and star rating.

¹ Programs with multiple classrooms had the option of completing a single program level SAS or multiple classroom level options. Psychometric analyses used the classroom-level SAS scores for programs with multiple classrooms and program level SAS records for programs with one SAS in the final SAS records.

² 275 programs and 344 SAS records were excluded from the psychometric and alternate score sample because of missing indicator-level SAS data. In the data file, zeros were used interchangeably to describe a score of 0 on an indicator and a missing value, making it impossible to differentiate between truly missing fields and incomplete SAS records. In the main analyses presented in the report, these 275 programs were excluded from all psychometric analyses. Sensitivity analyses that included these 275 programs resulted in similar patterns and interpretations of all psychometric and alternate score analyses.

License Type	1 STAR	2 STAR	3 STAR	4 STAR	5 STAR	Total	
Licensed Child Care Center	0	4	38	23	7	72	
Licensed Group Child Care Home	2	12	27	6	5	52	
Registered Family Child Care Home	12	8	27	5	6	58	
Total Programs	14	24	92	34	18	182	
Note. Data as of December 3, 2018							

Table 3-4. Number of Programs Participating by Program Type and Star Rating

Table 3-5. Number of Classrooms Participating by Program Type and Star Rating

License Type	1 STAR	2 STAR	3 STAR	4 STAR	5 STAR	Total	
Licensed Child Care Center	0	12	102	69	12	195	
Licensed Group Child Care Home	2	12	27	6	5	52	
Registered Family Child Care Home	12	8	27	5	7	59	
Total Classrooms	14	32	156	80	24	306	
Note. Data as of December 31, 2018							

e. Measures

Structured observations of center-based classrooms, family child care programs, and group child care programs can provide critical information about the quality of teaching practices, care routines, and experiences of participating children.

i. Extant Data

Extant program characteristic data available from Michigan was required by MDE and included program ID; name; address; ZIP code; e-mail and phone number; primary contact; initial QRIS quality level; community setting; program type (e.g., group- or family-based home care, center-based care, Head Start, GSRP, or for-profit program); enrollment size; and overall, subscale, and item-level SAS and Preschool Program Quality Assessment (PQA) data, as well as state-verified (and unverified) QRIS ratings.

ii. Observation Data

Classroom Assessment Scoring System (CLASS). The CLASS-Pre-K (Pianta, LaParo, & Hamre, 2008) is an observational assessment of the quality of teacher-child interactions. The CLASS domains include Emotional Support (i.e., positive climate, negative climate, teacher sensitivity, and regard for student perspectives), Instructional Support (i.e., quality of feedback and language modeling as well as concept development), and Classroom Organization (i.e., behavior

management, productivity, and instructional learning formats). Each dimension is rated from 1 to 7, with higher scores indicating higher quality. Data collectors observe classrooms for 20-minute cycles of observation, followed by 10 minutes of scoring. Classrooms were observed for 2–3 hours. Studies have found a link between CLASS domains and other measures of quality, such as the ERS (Early et al., 2006). The domain scores — Emotional Support, Classroom Organization, and Instructional Support — were used for analyses.

The Environment Rating Scales (ERS). The Environment Rating Scales are used to assess the global quality of child care programs. The ERS has three different versions appropriate for early childhood programs — the Early Childhood Environment Rating Scale-3 (ECERS-3; Harms, Clifford, & Cryer, 2015) for center-based preschool programs; the Family Child Care Environment Rating Scale-Revised (FCCERS-R; Harms & Clifford, 2007) for family child care programs; and the Infant/Toddler Environment Rating Scale-3 (ITERS-3; Harms, Cryer, Clifford, & Yazejian, 2017) for center-based programs serving infants and toddlers. These measures assess programs' structure, provisions for learning, and teaching and interactions. The various versions have 35–43 items, with subscales in the areas of (1) Space and Furnishing; (2) Personal Care Routines; (3) Listening and Talking; (4) Activities; (5) Interaction; (6) Program Structure; and (7) Parents and Staff (*Note*. This subscale was not collected because it could not be directly observed). Scores on the ERS can range from 1 to 7, with higher scores indicating higher quality. Studies have shown a relationship between the ERS and other indicators of program quality (Early et al., 2008) and child cognitive and social-emotional outcomes (Burchinal et al., 2008).

The total score for the ITERS-3, FCCERS-R, and ECERS-3 were used for analyses in addition to the subscale scores.

i. Expert Survey

Five national and local experts were consulted to review and rate the degree to which categories and indicators of the GSQ Self-Assessment System (SAS) represent components of quality. Experts had in-depth knowledge and experience in the following areas: QRIS validation and evaluation, continuous quality improvement, infant/toddler practices, family child care home practices, and birth-to-age-five program administration. Experts rated 49 items within the five categories of the SAS: (1) Staff and Professional Development; (2) Family and Community Partnership; (3) Administration and Management; (4) Environment; and (5) Curriculum and Instruction. Ratings ranged from 1 (not at all important) to 5 (very important) based on (a) how closely the item linked to program quality and (b) how closely the item linked to child outcomes. Experts were also asked to identify their top five indicators.

ii. Director Survey

Program directors and owners completed a survey with basic information about their program that observations cannot capture, including director/owner and educator education level and professional development, staff compensation policies and turnover, and perception about GSQ.

f. Data Collector Training and Quality Control

To ensure high-quality and reliable data, data collectors were trained as stipulated by protocols established by observation measure developers. For instance, all Classroom Assessment Scoring System (CLASS) observers were certified as reliable when they demonstrated 85% or higher (within one point) inter-rater reliability. Training was provided by certified trainers on the various quality measures. The Environment Rating Scales Institute trainer (ERSI-approved) provided training for the ERS tools and Teachstone-approved trainers provided training for CLASS. The Principal Investigators oversaw training on program quality measures and procedures.

Over 165 applications from around the state of Michigan were received from our job postings for data collectors. Seventy-five interviews were conducted in late June and July of 2018 with a focus on filling the positions with applicants spread across the 11 regions of the state. Applicants were required to have a minimum of an associate's degree (but preferably a bachelor's degree) in education or a related field, with our highest preference being early childhood education or child development. One to two years of experience in an early childhood care setting was also required and preference was given to applicants with experience administering program assessments. Data collector instrument training placements were based on related experiences with young children. In particular, applicants with infantand toddler-based experience were placed into ITERS-3 trainings. The data collector pool included students working on degrees beyond an associate's degree, retired early childhood teachers and administrators, independent early childhood contractors, and individuals working in part-time positions who were looking to increase their hours.

Data collector training for this project took place in August of 2018. In total, 11 different trainings took place; some of the training was conducted in Michigan and some in North Carolina. A total of 45 data collector slots were filled with 39 people, 6 data collectors having been cross-trained on a second instrument. On August 6, data collectors were trained on study procedures, data collector guidelines, and use of the data collector database.

The CLASS training took place in Ypsilanti, Michigan on August 7–8, 2018, and was conducted by a CLASS Affiliate Trainer. Once the data collectors completed the training sessions, they each completed the online reliability assessment. Of the 12 data collectors who trained on this tool, 10 were able to become reliable after one to three reliability attempts.

ECERS-3 data collectors each completed a 5-hour online introduction and training session provided by the Environment Rating Scales Institute (ERSI). Seven ECERS-3 data collectors then traveled to North Carolina for either a 2-, 3- or 5-day training provided by ERSI staff. Data collectors then returned to Michigan and were required to complete 2 consecutive days of reliability at a standard of 85% or better when matched with an anchor. Four ECERS-3 data collectors had 1 or 2 practice days in Michigan and were required to complete 2 consecutive days of reliability at a standard of 85% or better when matched with an anchor. Of the 11 data collectors who trained on the ECERS-3, 2 were able to serve as anchors (reliability scores above 88% across two days) and an additional 6 were able to become reliable at 85% or better. ITERS-3 data collectors each completed a 5-hour online introduction and training session provided by ERSI. Five data collectors then traveled to North Carolina for either a 2- or 4-day training provided by ERSI staff. Data collectors then returned to Michigan and were required to complete 2 consecutive days of reliability testing at a standard of 85% or better when matched with an anchor. One ITERS-3 data collector had 1 practice day in Michigan and 3 days of reliability testing to reach a standard of 85% or better when matched with an anchor. Of the 6 data collectors who trained on the ITERS-3, 2 were able to serve as anchors (reliability scores above 90% across 2 days) and the additional 4 were able to become reliable at 85% or better. Two ITERS-3 data collectors were not able to complete the training course.

FCCERS-R data collectors each completed a 5-hour online introduction and training session provided by ERSI. Three trainers traveled from North Carolina to conduct the data collector training and reliability. They each had reached the Gold Standard level in FCCERS-R and had collected data for projects in North Carolina. Six 4-day practice and reliability sessions were conducted in the greater Ypsilanti/Ann Arbor area and greater Detroit area. Of the 13 data collectors who trained on the FCCERS-R, all were able to become reliable at 85% or better. Four data collectors required an additional day of practice with a reliable partner.

IV. RESULTS

The findings of this study are divided into seven sections based on validation questions:

- 1.1 Are the QRIS quality components and standards the "right" ones?
- 1.2 Is the process of documenting and verifying each indicator yielding reliable and accurate results?
- 1.3 What is the relationship among the components and are they functioning as expected?
- 1.4 What are the most appropriate ways to combine measures of quality standards into summary ratings when examining the cut scores and combining rules?
- 1.5 What is the variation and pattern of program-level ratings within and across program types that ensures the ratings are functioning as intended?
- 2.1 Are there differences in GSQ ratings based on other measures of quality using the Environment Rating Scales (ERS) and Classroom Assessment Scoring System (CLASS) across program types?
- 2.2 How well do the ratings from the alternate cut points and rules distinguish different levels of quality?

1.1 Are the QRIS quality components and standards the "right" ones?

Experts' ratings of the SAS components indicate they are relevant for program quality, but less so for child outcomes.

To address this question, five national and local experts were consulted to review and rate the degree to which categories and indicators of the GSQ Self-Assessment System (SAS) represent components of quality. Experts rated 49 items within the five categories of the SAS: (1) Staff and Professional Development; (2) Family and Community Partnership; (3) Administration and Management; (4) Environment; and (5) Curriculum and Instruction. Ratings ranged from 1 (not at all important) to 5 (very important) based on (a) how closely the items linked to program quality and (b) how closely the items linked to child outcomes. Experts were also asked to identify their top 5 indicators.

Our analyses sought to determine the degree of agreement among the experts' ratings by match rate and intraclass correlations. High agreement would indicate that the experts agree that the categories and indicators of the SAS represent components of program quality and their importance for child outcomes, thereby providing evidence of construct validity.

Table 4-1 shows the results of the expert survey summarized by each one of the five SAS areas. We present summaries per rater as well as an overall summary across raters. First, a rater's evaluation of each area was estimated using the mean and median of the items contained in each of the five areas. For example, for the first area (Staff Qualifications and Professional Development) for program quality, the mean and median of the 16 corresponding items were computed for each one of the five raters. These means and medians are displayed in columns R1 to R5 in Table 4-1. For example, the mean for rater 1 was 2.8 and the median was 3 (see intersection of column R1 with rows 1 and 2, respectively).

Then descriptive statistics were computed across raters' evaluation of each area (means and medians per area) and are presented in Table 4-1: columns "Mean," "Median," "Min," and "Max." These statistics are interpreted as a summary across raters per each one of the five areas. For example, row 1 contains the mean, median, minimum, and maximum for the raters' mean of the 16 items in the Staff Qualifications and Professional Development area; while in row 2, the mean, median, minimum, and maximum values were computed across the five raters' medians displayed in row 2 and columns R1 to R5.

In general, the median of the medians would provide a more appropriate description, assuming that the expert ratings range in an ordinal scale using five values from "Not at all important" to "Very important"; however, if these ratings are assumed to be a continuum between "Not at all important" to "Very important" and the observed values are just arbitrary points of this continuum, then the mean of the means could be interpreted as another representation of the importance of the area as a component of quality.

Overall, the five experts' ratings of importance indicate that the five areas of the SAS are more relevant for program quality than for child outcomes. For, example, areas 1 (Staff Qualifications

and Professional Development) and 3 (Administration and Management) have median values of 4.0 for importance to program quality, but 3.0 when considering child outcomes.

			Across Raters Summary				F	Raters			
Usage	Area	Stat	Mean	Median	Min	Max	R1	R2	R3	R4	R5
	I. Staff Qualifications and Professional Development	Mean	3.8	4.2	2.8	4.7	2.8	4.7	4.2	4.2	3.0
		Median	4.0	4.0	3.0	5.0	3.0	5.0	4.0	4.0	3.0
	II. Family and	Mean	3.9	4.1	2.9	4.7	3.0	4.7	4.1	4.6	3.1
	Community Partnership	Median	3.8	4.0	3.0	5.0	3.0	5.0	4.0	5.0	4.0
Program	III. Administration and	Mean	3.9	4.1	3.0	4.7	3.2	4.7	4.2	4.7	3.0
Quality	Management	Median	3.8	4.0	2.0	5.0	3.0	5.0	4.0	5.0	3.0
		Mean	3.9	4.1	3.0	4.7	3.1	4.7	4.1	4.6	2.9
	IV. Environment	Median	4.2	4.0	3.0	5.0	3.0	5.0	4.0	5.0	2.0
	V. Curriculum and Instruction	Mean	3.9	4.2	3.0	4.7	3.1	4.7	4.1	4.5	3.0
		Median	3.8	4.0	3.0	5.0	3.0	5.0	4.0	4.0	3.0
	I. Staff Qualifications and Professional Development	Mean	3.6	4.3	2.1	4.7	2.6	4.7	4.3	4.4	2.1
		Median	3.6	3.0	2.0	5.0	3.0	5.0	5.0	4.0	2.0
	II. Family and Community Partnership	Mean	3.6	3.7	2.0	4.7	2.8	4.7	4.0	4.6	2.2
		Median	3.8	4.0	2.0	5.0	3.0	5.0	4.0	5.0	2.0
Child	III. Administration and	Mean	3.6	4.0	2.2	4.7	2.8	4.7	3.5	4.7	2.0
Outcomes	Management	Median	3.6	3.0	2.0	5.0	3.0	5.0	3.0	5.0	2.0
		Mean	3.7	4.2	2.2	4.7	2.9	4.7	3.7	4.6	2.0
	IV. Environment	Median	3.8	4.0	2.0	5.0	3.0	5.0	3.0	5.0	2.0
	V. Curriculum and	Mean	3.5	3.5	2.0	4.7	2.9	4.7	4.2	4.5	2.2
	Instruction	Median	3.8	4.0	2.0	5.0	3.0	5.0	5.0	4.0	2.0
Note. N = 5; Data as of December 31, 2018											

Table 4-1. Distribution of Relevance of Quality Indicators by Area and Across Raters

Experts do not agree which SAS indicators are important for program quality, but there is slightly more agreement for indicators that are important for child outcomes.

To examine the degree of agreement between the five experts, we estimated intraclass correlation (ICC) and Fleiss' Kappa for ordinal data with quadratic weights. (Linear weights provide much lower estimates.) The ICC tells the degree of agreement, expressed in percentage, between raters, on average, across all 49 indicators. The Fleiss' Kappa is an estimation of interrater reliability, similar to Cohen's Kappa, but specific for scenarios where there are more than

two raters and the rating scale is categorical. Larger values of ICC and Kappa imply greater agreement among raters. In examination of expert ratings of SAS indicators important for program quality, the ICC was .39 (CI95% [.07, .62]) with a Kappa Fleiss statistic for inter-rater reliability of .52 (CI95%[.45, .60]), indicating minimal to moderate agreement across the five experts about the importance of SAS indicators for program quality.

In examination of expert ratings of SAS indicators important for child outcomes, the results show moderate ICC of .70 (CI95%[.54, .82]) and a Kappa Fleiss statistic for inter-rater reliability of .24 (CI95%[.12, .35]), indicating slightly more agreement compared to the relevance for program based on the ICC differences; however, if we consider Fleiss statistics, there is not much agreement among raters for child outcomes.

Additionally, we calculated the proportion of raters who gave an indicator a value of 5 (highest quality) over the total number of raters. Overall, for both program quality and child outcomes, these rates are low. The average is around .43 for program quality and .41 for child outcomes. This means that, on average, raters did not agree on which SAS indicators are the most important for program quality and child outcomes. The median was .40 in both cases, with a range of 0 to .8 in both cases. Similarly, and in both children and programs, only 4 indicators out of 49 achieved a proportion of 0.8 (see Table 4-2).

Usage	Indicator
	Indicator 19: Center/program provides formal communication (i.e., parent/teacher conferences, home visits) to inform parents of children's developmental progress.
Child Outcomer	Indicator 23: Partnerships provide or connect families to appropriate comprehensive services.
child Outcomes	Indicator 32: Center/program is in a physical location that is free of environmental risks (e.g. lead, mercury, asbestos and indoor air pollutants).
	Indicator 48: Uses assessment to inform individual, small group, and whole group instruction and interaction.
	Indicator 20: Communication, education, and informational materials and opportunities for families are delivered in a way that meets their diverse needs (e.g., literacy level, language, cultural appropriateness, etc.).
Program Quality	Indicator 23: Partnerships provide or connect families to appropriate comprehensive services.
	Indicator 32: Center/program is in a physical location that is free of environmental risks (e.g. lead, mercury, asbestos and indoor air pollutants).
	Indicator 47: Uses child assessment results in parent-teacher conferences at least two times a year.

Table 4-2. Top 4 Indicators of SAS That Achieved Highest Rating for Child Outcomes and

 Program Quality

FINDINGS SUMMARY: National experts with extensive knowledge about QRIS, classroom quality, infant and toddler care, home-based programs, and program administration indicate

that the SAS indicators are more relevant for program quality than child outcomes. However, they did not agree which SAS indicators were more relevant for program quality and child outcomes. In general, they agree that the categories of the SAS — staff qualifications, family and community partnership, administration management, environment, and curriculum and instruction — matter.

1.2 Is the process of documenting and verifying each indicator yielding reliable and accurate results?

The GSQ ratings are reliable based on an independent scoring of the standards and associated program documentation.

Another aspect of this validation study was to examine the reliability of the system to ensure that GSQ is not biased or flawed in its process of rating programs based on published standards; that is, we sought to examine whether the ratings were reliable based on the components and standards of the system and the documentation of these standards. To examine this, we selected 60 programs across different star ratings and types (see Table 4-3). By the time of review, however, 2 of the group child care homes were no longer licensed and eligible for this sub-study.

1 STAR	2 STAR	3 STAR	4 STAR	5 STAR	Total
	8	8	7	7	30
2	4	2	4	3	15
4	2	4	2	3	15
6	14	14	13	13	60
	1 STAR 2 4 6	1 STAR 2 STAR 8 2 4 4 2 6 14	1 STAR 2 STAR 3 STAR 8 8 2 4 2 4 2 4 6 14 14	1 STAR 2 STAR 3 STAR 4 STAR 8 8 7 2 4 2 4 4 2 4 2 6 14 14 13	1 STAR 2 STAR 3 STAR 4 STAR 5 STAR 8 8 7 7 2 4 2 4 3 4 2 4 2 3 6 14 14 13 13

Table 4-3. Program Type and Star Rating for Reliability Sub-Study

Note. Michigan Data as of June 14, 2018

Source. Michigan Department of Education, Office of Great Start

A member of the research team reviewed the standards for each type of program and the documentations provided by the programs and PQA scores from ECIC (GSQ implementation partner), when relevant. This research team member assigned points for each component of the SAS based on documentation provided by ECIC. Summary statistics on Table 4-4 show that HighScope's independent rating was almost identical at the points and rating level. For example, for the 30 centers that were reviewed, the average star rating was 3.4, which was the same for the official MDE rating. Furthermore, as shown on Table 4-5, there was high correlation between HighScope's star rating and MDE's star ratings, with correlations ranging from .78 to .98.

			MDE		High	Scope
License type	Area	Ν	Mean	Std	Mean	Std
	I. Staff Qualifications and Professional Development	30	8.8	4.1	8.5	4.3
	II. Family and Community Partnership	30	6.5	2.0	6.7	1.6
Licensed Child Care Center	III. Administration and Management	30	5.3	1.4	5.4	1.2
	IV. Environment	30	5.6	2.0	5.7	1.8
	V. Curriculum and Instruction	30	8.0	3.4	8.3	3.3
	Total	30	34.2	10.0	34.7	10.0
	Rating	30	3.4	1.2	3.4	1.1
	I. Staff Qualifications and Professional Development	15	5.9	3.6	4.9	4.2
	II. Family and Community Partnership	15	5.7	1.8	5.7	2.0
Registered Family Child Care Home	III. Administration and Management	15	5.1	1.3	4.8	1.5
	IV. Environment	15	5.1	1.7	5.3	1.4
	V. Curriculum and Instruction	15	5.9	4.6	5.5	4.9
	Total	15	27.7	11.5	26.3	12.3
	Rating	15	2.7	1.5	2.7	1.4
	I. Staff Qualifications and Professional Development	13	7.8	4.7	7.8	4.9
	II. Family and Community Partnership	13	5.7	2.4	5.8	2.2
Licensed Group Child Care Home	III. Administration and Management	13	4.6	2.4	4.9	2.3
	IV. Environment	13	5.8	1.3	5.8	1.3
	V. Curriculum and Instruction	13	5.7	4.7	5.5	3.9
	Total	13	29.6	12.2	30.0	12.0
	Rating	13	3.0	1.4	3.2	1.4
<i>Note.</i> n = 55						

Table 4-4. Summary Statistics for SAS Validation Ratings by License Type per Area

HS = HighScope MDE Data as of June 14, 2018

Туре	Correlation	Lower	Upper	P-value				
Center	0.785	0.5930	0.893	0.000				
Group Child Care Homes	0.977	0.922	0.993	0.000				
Family Child Care Homes	0.779	0.444	0.923	0.001				
Total	0.851	0.760	0.909	0.000				
Note. n = 58 Correlation for Family Homes with assistance was 1								

Table 4-5. Correlation Matrix Between HighScope Rating and MDE Rating of the Same Programs

FINDINGS SUMMARY: An independent rating of 60 programs using the MDE standards about program quality for center- and home-based programs resulted in the same official ratings. Specifically, there was more congruence between the independent rater for group child care homes compared to centers and family child care homes.

1.3 What is the relationship among the components and are they functioning as expected? That is, how do the SAS and the PQA correlate to each other and to the overall rating of quality?

Two-thirds of programs' self-ratings matched the ratings they received through independent observations of quality. However, about a third of programs rated themselves higher than confirmed through independent observations of quality.

As part of the GSQ, if a program is interested in a final published score of 4 or 5 stars, staff must first complete the SAS with a self-rating of either 4 or 5 stars. Next, they must apply for and complete an independent observation of quality to confirm the 4- or 5-star rating. Following the state-administered independent observation (PQA), the program's self-ratings are confirmed as a 4- or 5-star rating, or may decrease to a 3- or 4-star rating if the independent observation of self-rating.

As of June 14, 2018, 237 programs had both a self-rating on the SAS and an independent observation of quality on the PQA. Of these programs, the independent observation of quality confirmed the self-rating on the SAS for 158 programs (67%). These programs then received their published GSQ rating as this score. Observations were not always concordant with self-ratings. Seventy-nine programs (33%) received a rating on the independent observation that was one or more levels lower than the self-assessment rating (see Table 4-6). These programs received a final published GSQ rating representing their independent observation of quality. No programs had a rating on the independent observation that was higher than the self-assessment rating.

	Rating on Independent Observation of Quality						
Self-Rating	Level 3 (<i>n</i> = 27)	Level 4 (<i>n</i> = 142)	Level 5 (<i>n</i> = 68)				
1000 4 (n - 106)	23	83	0				
Level 4 ($n = 106$)	(9.7%)	(35.0%)	(0.0%)				
$\lfloor \alpha_{n} \alpha \rfloor \left[\Gamma \left(n - 1 \right) \right]$	9	90	32				
Level 5 ($n = 131$) -	(3.8%)	(38.0%)	(13.5%)				

Table 4-6. Comparison of Self-Rating and Independent Observation of Quality

n = 237.

Note. Data in boldface indicate that self-assessment ratings and ratings on the independent observations of quality were the same. Data are current as of June 14, 2018.

Source: Calculations based on data provided by the Michigan Department of Education, Office of Great Start.

To understand how the SAS and PQA were associated with each other, we also conducted correlational analyses for the set of programs with both an SAS and a PQA observation. The correlation between the SAS and PQA was positive and statistically significant, but lower than expected (Spearman's $\rho = 0.38$; $\rho < .05$). This low association may suggest that the two instruments measure different aspects of program or classroom quality, or that the SAS or PQA do not measure quality well for programs with 3- through 5-star ratings.

Both the self-report rating and independent observations of quality are related to the final GSQ rating with the GSQ rating more strongly related to the final self-report rating.

All three measures were positively and significantly correlated with one another (the SAS, PQA, and final published GSQ star ratings), but to varying degrees (see Table 4-7). Final SAS ratings and published GSQ ratings were nearly perfectly correlated (Spearman's $\rho = 0.95$; p < .05). The PQA was moderately correlated with the published star ratings (Spearman's $\rho = 0.67$; p < .05; see Table 4-7).

Table 4-7. Correlations Among the Final SAS Rating	, Observed PQA Score, and Published GSQ
Rating	

	Final SAS Rating	Observed PQA Score	Published GSQ Rating
Final SAS Rating	1.00		
Observed PQA Score	0.38*	1.00	
Published GSQ Rating	0.95*	0.67*	1.00

n = 237.

Note. All correlations are Spearman rho; * indicates significance at the alpha = 0.05 level. SAS is Self-Assessment Survey. PQA is Program Quality Assessment. QSQ is Great Start to Quality. Data are current as of June 14, 2018. *Source:* Calculations based on data provided by the Michigan Department of Education, Office of Great Start.

To what extent do the SAS subdomains measure the underlying aspects of quality as intended?

In this section, we focus specifically on the self-assessment rating and the extent to which the SAS subdomains are measuring the underlying aspects of quality as intended. Given the contribution of the SAS to a program's final GSQ rating, it is important that the measure is valid and reliable. The expectation is that the SAS measures meaningful aspects of program quality. This also means that one can assume that differences in program scores on the SAS reflect meaningful differences in program quality, rather than some other factor. To test these assumptions, we conducted a series of psychometric analyses to learn how 38 indicators — including those indicators that overlapped across the three versions of the SAS (i.e., center- based protocol, home-based with assistants protocol, and home-based without assistants protocol) — grouped together to form subdomains of program quality. First, we conducted exploratory factor analyses to understand which items measured the same factors of quality. Next, we conducted Rasch analyses to examine which indicators of quality were the easiest or most difficult for

participants to achieve, and which items give us the most information about program quality. Key findings are presented in this section, and more detail on these analyses is provided in Appendices A and B.

Within the SAS, all five domains are significantly correlated with one another and the final SAS rating.

We began the analysis of the SAS by examining how the subdomains of the SAS functioned together to achieve the final SAS rating. The SAS protocols were designed to measure five distinct aspects of program quality: (1) Staff Qualifications and Professional Development; (2) Family and Community Partnerships; (3) Administration and Management; (4) Environment; and (5) Curriculum and Instruction. All five subdomains on the SAS were positively and significantly correlated to the final SAS rating (see Table 4-8). These correlations vary in strength, however. The items on the Staff Qualifications and Professional Development subdomain and the Curriculum and Instruction subdomain had the strongest correlations with the final SAS rating (Spearman's $\rho > 0.80$). This suggests that these two subdomains are most closely related to the overall SAS scores, or that these items "drive" the overall SAS score. It also may suggest that these two subdomains may be redundant or that they may, in fact, measure the same underlying aspects of quality. Administration and Management had the lowest correlations, with a Spearman's $\rho = 0.44$; p < .05, suggesting that it contributes the least to the overall SAS score.

Finally, we examined the correlation of each indicator to its assigned program quality domain score. While we would expect some correlation between subdomains because programs that perform well in one area are likely to perform well in another, higher correlations between subdomains may indicate that items within these subdomains are picking up on similar aspects of quality rather than distinct aspects of quality, as intended. We see moderately high correlations between the Curriculum and Instruction subdomain and the Staff Qualifications and Professional Development subdomain (Spearman's $\rho = 0.72$; p < .05). We also see strong correlations between the Curriculum and Instruction subdomain and the Family and Community Partnerships subdomain (Spearman's $\rho = 0.67$; p < .05). These high correlations may suggest that the items in these domains overlap and measure the same underlying construct of program quality. This finding is further supported by the exploratory factor analyses.

SAS Subdomain	Overall Self- Assessment Rating (1, 2, 3, 4, or 5)	Administration and Management	Curriculum and Instruction	Environment	Family and Community Partnerships
Administration and Management	0.44***				
Curriculum and Instruction	0.89***	0.38***			
Environment	0.58***	0.19***	0.47***		
Family and Community Partnerships	0.72***	0.38***	0.67***	0.41***	
Staff Qualifications and Professional Development	0.84***	0.30***	0.72***	0.43***	0.53***
*** <i>p</i> < .0001 N = 3,821					

Table 4-8. Correlation of SAS Sub-Domains Scores With Each Other and Self-Assessment Ratings

Note. All correlations are Spearman rho. Data are current as of June 14, 2018.

Source: Calculations based on data provided by the Michigan Department of Education, Office of Great Start.

Some "easy items" are affecting the subsequent analyses.

To inform the exploratory factor analyses and Rasch analyses, the research team examined response patterns on the SAS. Responses to 17 items on the SAS were highly skewed toward a "yes" response, with over 90% of programs self-rating as having this aspect of quality (Table A1 in Appendix). For example, over 95% of SAS respondents reported that they communicate informally with parents, that they have written personnel policies and procedures, and that their program is free of environmental risks. These highly skewed items were particularly problematic for the psychometric analyses. With the presence of these highly skewed items, the statistical analyses that relied on a covariance or correlation matrix could not correctly compute standard errors. Therefore, although the research team continued to conduct preliminary analyses on exploring the factor structures based on the extant data, the overall results from factor analyses may not be trustworthy.

In the sections that follow, we present the preliminary analyses from the psychometric analyses. As we move forward with the next stages of the extant data analysis, we will examine ways in which we can improve the measures (e.g., removing items that are not contributing to the underlying aspects of quality, rethinking how items are grouped and scored, and making recommendations for adding or improving items to strengthen the intended survey structure). These findings are presented for discussion purposes only, and interpretations of the results should be made with extreme caution.

Factor analyses revealed little evidence that the SAS items measure five distinct aspects of program quality as intended.

Given the concerns about the lack of variability in the self-ratings, we performed initial psychometric analyses to learn how the SAS items function. The purpose of the exploratory factor analyses (EFAs) was to test a series of assumptions about how items and sets of items correlate with one another and how the items function together. If the SAS was functioning as intended, we would see five unique factors that correspond with the original five subdomains of the SAS: (1) Staff Qualifications and Professional Development; (2) Family and Community Partnerships; (3) Administration and Management; (4) Environment; and (5) Curriculum and Instruction. However, the EFAs suggested that SAS items may, in fact, measure the same underlying aspects of program quality. Specifically, we found that a five-factor model did not fit the data. Instead, we found a possible two-factor or three-factor solution, including the following subdomains:

- *General Staff Qualifications.* This domain includes items from the original Staff Qualifications and Professional Development domain. Items include director's educational background and staff's educational background.
- General Program Operation. This domain includes items primarily from the original Family and Community Collaboration domain, the Environment domain, and the Curriculum and Instruction domain. Example items include "Center provides parenting education opportunities" and "Center is in a physical location that is free of environmental risks (e.g., lead, mercury, asbestos, and indoor air pollutants)."
- *Child Assessment and Screening.* This domain includes items primarily from the original Curriculum and Instruction domain that ask about how children are assessed.

The other two factors that best fit the data were uninterpretable. One factor has no factor loadings above 0.40 (the recommended cutoff for factor loadings), and another had a mix of items that did not measure one underlying factor of quality. The factor loadings for all five possible factors are presented in Table A3 in Appendix A.

The correlations among the General Staff Qualifications domain, the General Program Operation domain, and the Child Assessment and Screening domain were 0.18 (Qualification and Operation), 0.09 (Qualification and Screening), and 0.64 (Operation and Screening), all with p < .05. These correlations are lower than the correlations found between each pairing of the five factors hypothesized by the SAS developers. The moderate correlation indicates that there was sufficient distinction among staff qualification, program operation, and screening aspects of program quality. However, this three-factor solution was only an "emerging" theme and was distant from a final solution. The research team would not endorse this three-factor model because the analyses in general did not reveal any meaningful measurement structure based on the current data. Next steps include re-running these factor analyses with fewer items to identify a more ideal factor structure.

The Rasch analyses also revealed little evidence that the SAS items measure five distinct aspects of program quality as intended.

Next, the research team conducted Rasch analyses to understand how well the items within each distinct aspect of program quality on the SAS functioned. For each subdomain on the SAS, the Rasch analyses provided estimates of internal consistency, or the degree to which items measure the same underlying construct. The Rasch analyses also identified which items are easy (e.g., all programs have this aspect of quality) or difficult (e.g., few programs participating in GSQ have this aspect of quality). When items function well together, clear patterns emerge from which we can derive the difficulty³ of each item within a domain of quality, as well as the likelihood that a program will meet a given indicator (e.g., there will be "higher" and "lower" rated programs). For example, in a well-functioning measure, we see survey respondents consistently agreeing more strongly with one item and less strongly with another item. With multiple items within a domain, we can then estimate each program's "score" on the domain of quality based on the responses of survey participants relative to those of other respondents.

Before calculating the overall "score," we examined the internal consistency and reliability of each SAS domain to determine whether the domains were functioning as intended. First, we examined two measures of reliability — person separation reliability and Cronbach's alpha both of which measure how consistently respondents answered each question relative to their other responses as well as to one another. For a well-functioning measure, these two measures typically are similar. For both measures, reliability ranges from 0 (no reliability) to 1 (perfect reliability). Scores above 0.70 are considered good and scores above 0.80 even better. For the SAS, the alphas were consistently above 0.80 (0.84 for centers, 0.81 for homes, and 0.92 for the combined sample), suggesting good internal consistency across all items. However, the person separation reliability — how well we can distinguish the high versus low raters — for the five subdomains was much lower (see Table 4-9). For four of the SAS subdomains, the person separation reliability was 0.00 or 0.29. This indicates that the items within these domains are not sensitive enough to measure differences between high- and low-quality respondents, which was the case on 17 items on the SAS. The person separation reliability and measure of internal consistency were higher for the Staff Qualifications and Professional Development domain and the Curriculum and Instruction domain, indicating that the items within these SAS subdomains may reliably measure differences in program quality as intended.

³ In this case, difficulty is related to how many programs achieved each item on the SAS. *Easy* items are those items on which many or all programs self-rated as having this aspect of program quality. *Difficult* items are those on which few or no programs self-rated as having this aspect of program quality.

Program Quality Domain	Number of Items	Person Separation Reliability	Internal Consistency (Cronbach's Alpha)	Range of Item Fit	Point- Correlation Range
Staff Qualifications and Professional Development	7	0.69	0.72	0.70–-1.36	0.32–0.79
Family and Community Partnerships	9	0.00 ^a	0.85	0.66–2.16	0.26–0.71
Administration and Management Homes ^b	4	0.00 ^c	0	0.99–1.01	0.39–99.4
Administration and Management for Centers	5	0.29	1	0.86–1.14	0.43–0.69
Environment	7	0	0.34	0.81–1.27	0.10-0.54
Curriculum and Instruction	11	0.72	0.92	0.31–9.90	0.14–0.87

Table 4-9. Person Separation Reliability, Internal Consistency, Range of Item Fit, and Point-Correlation Range of SAS Program Quality Domains

Note. The sample includes possible SAS items across centers and homes. Data are current as of June 14, 2018.

^a55% of all respondents received the maximum number of points on this scale, therefore decreasing the ability of the instrument to accurately measure differences in quality between respondents. For the 1,957 respondents without the maximum score, the person separation reliability was low (reliability = 0.10) due to total scores at the higher end of the distribution.

^bThe Administration and Management items were run separately because none of these items overlapped. There was a total of N = 1,441 for home-based SAS records and N = 3,491 for center-based SAS records.

^c73% of all respondents received the maximum number of points on this scale, therefore decreasing the ability to accurately measure differences in quality between respondents. For two items, all but three respondents answered "Yes."

Source: Calculations based on data provided by the Michigan Department of Education, Office of Great Start.
In addition to the measures of reliability, the research team examined the item fit and the correlations between each item and the overall measure. This analysis provides a measure of "outfit" for each item on the SAS. *Outfit* measures the extent to which the observed ratings align with expected ratings based on the ratings on all other items (Bond & Fox, 2007). An outfit of 1.00 indicates a perfect fit of an item with the domain, and an outfit between 0.50 and 1.50 is considered good fitting. For three SAS subdomains, the items fit well with the overall measure — Staff Qualifications and Professional Development, Administration and Management for Centers, and Environment. For one subdomain, Administration and Management for Homes, the item fit was within the expected range; however, because of the skewed nature of the data, in which 90% or more of the early childhood education programs rated themselves as having these indicators of quality, this subdomain includes problematic items. Finally, for the Family and Community Partnerships domain and the Curriculum and Instruction domain, we saw some items that did not have good fit. Although beyond the scope of this current analysis, this finding indicates that we may want to remove some items from these constructs to improve the overall quality of the measure.

Finally, we examined the correlation of each item to the overall SAS score. The point correlations range from 0.00 (no correlations) to 1.00 (perfect correlations), and we expect items that fit well to correlate well with the overall measure in general. As shown in Table 4-9, there are items with low correlations (less than .40); however, these items likely would be removed or improved through future measure development (e.g., removal of poor-fitting items or introduction of new items to a domain).

The Rasch analysis also suggests that the SAS items are misaligned to program reports of quality.

In addition to looking at the items within the intended structure, we ran a Rasch analysis on a single-factor structure (i.e., included all available items combined across program types) to learn more about how the SAS items function (i.e., how difficult or easy they are) relative to self-reports of program quality. In a well-performing measure, we would expect to see a range of easy to difficult items that are well aligned with differences in reported quality across the respondents. As shown in Figure 4-1, items typically were too easy relative to program reports of quality. Given the response patterns observed in this sample, we expect all participants to receive points for the easiest items (e.g., items with a scale score < -2.5 logits), whereas at the higher end of the scale, we do not have any items difficult enough to detect differences in the highest performers (e.g., respondents with a scale score > 3.0 logits). This suggests that the SAS could be improved as a measure through the addition of more difficult items, or it may suggest that programs generally are of high quality across the state and are meeting the indicators of program quality in the GSQ.



Figure 4-1. Single-Factor Rasch Analysis Relative to Self-Assessment of Program Quality

Given the findings in the EFA and Rasch analyses, the SAS as currently designed may not be measuring the intended five subdomains of quality as intended. In the next series of analyses for research question 1.4 (next section), we will streamline the SAS and cut items that are not a good fit. We will also re-weight indicators to identify an alternate soring approach that yields more meaningful differences for programs and integrates their CLASS and ERS scores.

FINDINGS SUMMARY: To understand whether the Self-Assessment System (SAS) was functioning as expected, several different analyses were conducted including correlation, factor analyses, and Rasch analyses. Examination of SAS and GSQ ratings and independent observations found that over two-thirds of programs' self-ratings matched the ratings they received through independent observations of quality. However, about a third of programs rated themselves higher than confirmed through independent observations of quality. There were no programs that rated themselves lower. Correlation analyses found that both the self-report and independent observations of quality are related to the final GSQ rating with the GSQ rating more strongly related to the SAS than the independent observation of quality. This indicates that the SAS is a big factor in deciding final ratings. Correlation analyses also showed that all five domains within the SAS are significantly correlated with one another and the final SAS rating. The overall SAS rating is also highly correlated with curriculum and instruction, staff qualifications, and family and community partnership. Exploratory factor and Rasch analyses found that there were not five meaningful factors that aligned with the SAS categories, likely due to the many easy items where programs self-rated as having this aspect of quality.

1.4 What are the most appropriate ways to combine measures of quality standards into summary ratings when examining the cut scores and combining rules?

In this section, we examined two alternate scoring approaches to address RQ 1.4: Examining the cut scores and combining rules to determine the most appropriate ways to combine measures of *auality standards into summary ratings*. To address this research question, with input from MDE, we developed two alternate scoring approaches: 1) The "streamlined" approach and 2) the "Few and Mighty" approach. These alternate scoring approaches focused primarily on changes to the SAS, given MDE's decision to allow programs to select the PQA or CLASS as their classroom observation measure for GSRP (Michigan's Pre-K program) beginning in 2019. The recommended alternate scoring approaches for the SAS were based on all extant analyses, a review of the literature about aspects of program quality that are most closely related to child outcomes, feedback from the expert panel members on each indicator, and center directors' input on the most important domains of quality collected via surveys. In the first alternate scoring approach, the "streamlined" approach, we reduced the number of indicators to focus on those indicators with the most variability across programs. In the second alternate scoring approach, the "Few and Mighty" approach, we made a more severe reduction in the total number of indicators to focus on a set of indicators with high variation and grounded in the literature describing what matters most for quality in early childhood programming.

In addition to reducing the number of SAS indicators for the two alternate scoring approaches, we reduced the number of quality tiers from five to three levels. This recommendation is based on national validation studies that often find meaningful differences between the highest and lowest ratings, but not necessarily among five distinct levels of quality (Tout et al., 2017). The three levels of quality we recommend are:

- Level 1 Improvement Planning, where programs that met all licensing requirements are beginning the process of planning for improvement;
- Level 2 Improvement Progress, for programs in the process of improvement based on their goals and activities outlined in the Level-1 improvement planning stage; and
- Level 3 High Quality, for programs that met high-quality standards, as verified through an observation process. In addition, programs that rate at Level 3 must demonstrate participation in ongoing improvement efforts reflecting continuous professional growth and development of all program staff.

Using these alternate levels, we created a series of simulated scores for each program using their SAS scores, yielding new alternate ratings ranging from levels 1 to 3. We then examined the distribution of the simulated SAS ratings for both alternate scoring approaches and compared these scores with the original SAS indicators to determine the extent to which programs changed (i.e., moved to a higher or lower rating) for each alternative scoring approach.

The streamlined approach

The purpose of the streamlined approach was to reduce the number of indicators on the SAS to: (1) create a survey with stronger psychometric properties, and (2) reduce the time required by programs to complete the survey. In this approach, we removed indicators with limited variability. We also removed indicators that overlapped with current licensing requirements, because all programs in GSQ must already be licensed child care centers or homes and, therefore, have already met these indicators. We also reviewed the expert panel review to identify indicators that should remain in the SAS. We also incorporated feedback from center directors about the most important domains in the SAS, from their perspective. Finally, we incorporated the aspects of program quality that were most related to child outcomes in the early childhood literature. Figure 4-2 shows a comparison of the domains and number of indicators for the existing and streamlined SAS. As shown, the streamlined version includes 28 indicators for centers and 23 indicators for homes from the SAS (Figure 4-2 and Table 4-10). Figure 4-3 shows the placement of individual indicators by quality domain in the final streamlined approach.



Figure 4-2. Domains and Number of Indicators in the Original and Streamlined SAS

Note. NA means that indicators were not relevant for homes and therefore not included in the streamlined SAS for homes.



Figure 4-3. Domains and Number of Indicators in the Original and Streamlined SAS

In addition, the center and home streamlined versions of the SAS differ slightly. For example, centers have more staff qualifications indicators that correspond with the education levels of center directors, teachers, and assistants, while homes only have one or two indicators depending on if the provider has an assistant or not. Also, a sixth scale was added for centers and titled "Staff Retention." This scale includes three indicators from the prior administration and management scale and measures aspects of the center's management structure that would support teacher retention and continuity of care for infants, toddlers, and children. The total possible number of points on the streamlined SAS is 32 for centers and 27 for homes. See Table 4-10 for a list of all indicators in the original SAS and if they were cut or kept for the home and center-based versions of the streamlined approach.

Subdomain of			Changes Streamline	Made for d Approach
the SAS	Indicator	Item Wording	Centers	Homes
Director education and training	DIRECTOR1	CDA or Montessori credential including a minimum of 18 semester hours in Early Childhood Education/Child Development and 960 hours of experience OR Bachelor's degree or higher in an unrelated field with a minimum of 18 semester hours in Early Childhood Education/Child Development and 960 hours of experience		Revised to 1 item for homes
	DIRECTOR2	Associate's degree in Early Childhood Education/Child Development or child- related field including a minimum of 18 semester hours Early Childhood Education/Child Development and 480 hours of experience OR 60 semester hours in a program leading to a Bachelor's degree in Early Childhood Education/Child Development or child-related field with at least 24 semester hours in Early Childhood Education/Child Development and 480 hours of experience	Revised to 1 item for centers	
	DIRECTOR3	Bachelor's degree or higher in a child-related field including a minimum of 18 semester hours in Early Childhood Education/Child Development and 480 hours of experience OR Bachelor's degree or higher in any field with 30 semester hours in Early Childhood Education/Child Development and 480 hours of experience		
	DIRECTOR4	Bachelor's degree or higher with a major in Early Childhood Education/Child Development and 2 credits in child care administration		
Staff lead provider/ educator/teacher qualifications	STAFF1	At least 50% of classrooms have lead providers/educators/teachers with, at a minimum, a CDA OR Montessori credential as appropriate		Cut
	STAFF2	100% of classrooms have lead providers/educators/teachers with, at a minimum, a CDA OR Montessori credential		
	STAFF3	At least 50% of classrooms have lead providers/educators/teachers with at a minimum an Associate's degree in Early Childhood Education/Child Development or child-related field OR 60 semester hours in a program leading to a Bachelor's degree in Early Childhood Education/Child Development or child-related field with at least 24 semester hours in Early Childhood Education/Child Development		
	STAFF4	100% of classrooms have lead providers/educators/teachers with at least an Associate's degree in Early Childhood Education/Child Development or child- related field including a minimum of 18 semester hours in Early Childhood Education/Child Development OR 60 semester hours in a program leading to a Bachelor's degree in Early Childhood Education/ Child Development or child-related field with at least 24 semester hours in Early Childhood	Revised to 1 item for centers	
		Education/Child Development		
	STAFF5	At least 50% of classrooms have lead providers/educators/teachers with at least a Bachelor's degree in Early Childhood Education/Child Development or child-related field		
	STAFF6	100% of classrooms have lead providers/educators/teachers with at least a Bachelor's degree in Early Childhood Education/Child Development or child- related field		

Table 4-10. Indicators Included in the Streamlined Approach Versus the Original SAS

Subdomain of			Changes Made for Streamlined Approach	
the SAS	Indicator	Item Wording	Centers	Homes
	STAFF7	At least 50% of assistants have at a minimum a CDA or Montessori credential appropriate to age served OR 100% of assistants have completed at least one post-secondary course in Early Childhood Education/Child Development OR 20 hours of community/academic training aligned with the Core Knowledge Core Competencies		
	STAFF8	100% of assistants have at a minimum a CDA OR Montessori credential		
Staff assistant provider/ educator/teacher qualifications	STAFF9	At least 50% of assistants have at a minimum an Associate's degree in Early Childhood Education/ Child Development OR A child-related field including a minimum of 18 semester hours in Early Childhood Education/Child Development OR 60 semester hours in a program leading to a Bachelor's degree in Early Childhood Education/Child Development or child-related field with at least 24 semester hours in Early Childhood Education/Child Development	Cut	Cut
	STAFF10	100% of assistants have at a minimum an Associate's degree in Early Childhood Education/Child Development OR A child-related field including a minimum of 18 semester hours in Early Childhood Education/Child Development OR 60 semester hours in a program leading to a Bachelor's degree in Early Childhood Education/Child Development or child-related field with at least 24 semester hours in Early Childhood Education/Child Development		
	PD1	Director and all program staff complete at least 24 clock hours of professional development annually	Кеер	Keep (X hours for homes)
Professional development	PD2	Annual professional development training attended by all staff includes at least 3 hours focused on cultural competence OR inclusive practices, related to serving children with special needs or disabilities, as well as teaching diverse children and supporting diverse children and their families	Кеер	Кеер
Other professional training	PD3	Director has a graduate degree in Early Childhood or Child Development or a related field OR Program works at least monthly with an Early Childhood Specialist with a graduate degree in Early Childhood or Child Development or a related field	Integrated into the Center Director Item above	Integrated into the Home Provider Item above
	PD4	Center develops Quality Improvement Plan designed to improve quality in staff qualifications and progress is monitored by a Quality Improvement Consultant	Кеер	Кеер
	FAMILY1	Center provides parenting education opportunities	Cut	Cut
	FAMILY2	Classroom staff engages in informal communication with parents	Cut	Cut
Family partnerships and family strengthening	FAMILY3	Center provides formal communication (i.e., parent/teacher conferences, home visits) to inform parents of children's developmental progress	Кеер	Кеер
	FAMILY4	Communication, education, and informational materials and opportunities for families are delivered in a way that meets their diverse needs (e.g., literacy level, language, cultural appropriateness, etc.)	Кеер	Кеер
	FAMILY5	Center offers opportunities for parents to participate in program governance	Cut	Cut
	FAMILY6	Center provides opportunities for parents to participate in education inside and outside the classroom	Cut	Cut
Community partnerships	COMMUNITY1	Partnerships to provide or connect families to appropriate comprehensive services	Кеер	Кеер
	COMMUNITY2	Partnerships that take basic steps to facilitate children's transition between and among programs, agencies and schools	Кеер	Кеер
	COMMUNITY3	Participation in community associations	Cut	Cut

Subdomain of			Changes Made for Streamlined Approach	
the SAS	Indicator	Item Wording	Centers	Homes
	ADMIN1	Written personnel policies and procedures	Not directly	NA
	ADMIN2	Evidence of staff evaluations and individual professional development plans for each staff member	related to child	
Administration and	ADMIN3	A documented, graduated salary scale for staff that takes into account education and experience	outcomes, keep in a	
management	ADMIN4	A flexible benefit plan that may include health, tuition assistance, etc., for staff	staff retention	
	ADMIN5	Paid leave time for full-time employees which may include holiday, vacation, educational leave, and/or sick time	centers	
Physical environment	PHYSICAL1	Center is in a physical location that is free of environmental risks (e.g., lead, mercury, asbestos and indoor air pollutants)	Cut (licensing requirement)	Cut (licensing requirement)
Ratios	RATIOS1	Center demonstrates that it has smaller group size and better teacher:child ratio than required by licensing	Кеер	Кеер
Healthy environment	HEALTH1	Center is participating in the CACFP in good standing and has a written nutrition plan: OR follow seasonal menu guidelines that meet the CACFP meal pattern requirements and have a written nutrition plan; OR for programs that serve a snack only-follow seasonal menu guidelines that meet CACFP meal pattern requirements; OR for programs that do not provide food-provide nutrition information to families if families provide meals from home	Cut (licensing requirement)	Cut (licensing requirement)
	HEALTH2	30 minutes of every 3 hours dedicated to active outdoor time, with appropriate indoor physical activities available when weather prohibits outdoor play	Кеер	Кеер
	HEALTH3	Provisions for reviewing and updating health records according to the most recent Early, Periodic Screening, Diagnosis and Treatment (EPSDT) schedule for infants, and reviewing and updating records for toddlers and older children annually	Keep	Кеер
	HEALTH4	A process for observing each child's health and development on a daily basis and communicating observations to the child's family, other provider/educators and to specialized staff, with recommendations for family to seek medical opinions as necessary	Кеер	Кеер
	HEALTH5	A regular oral care routine, including tooth brushing and/or gum wiping (for infants) at least once per day	Кеер	Кеер
	CURRICULUM1	A statement of educational and developmental priorities for the children	Cut	Cut
	CURRICULUM2	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	Combined with outdoor play item	Combined with outdoor play item
Curriculum	CURRICULUM3	An approved curriculum	Кеер	Кеер
	CURRICULUM4	A written plan for integrating policies, procedures and practices that reflects a respect and valuing of children's culture and demonstrates cultural competence	Кеер	Кеер
	CURRICULUM5	A written plan for serving children with special needs	Кеер	Кеер
Screening and assessment	SCREENING1	Staff discusses anecdotal notes/observations as a basis for working/teaching with each child	Кеер	Кеер
	SCREENING2	Complete annual developmental screening on each child	Кеер	Кеер
	SCREENING3	Uses an approved child assessment tool at least two times a year	Кеер	Кеер
	SCREENING4	Uses a child assessment results in parent-teacher conferences at least two times a year	Кеер	Кеер
	SCREENING5	Uses assessment to inform individual, small group and whole group instruction and interaction	Кеер	Кеер
Consistent caregiving	CONSISTENT1	Child care/Preschool center can demonstrate that it structures and schedules staff such that each child has a consistent team of provider/educators and peers over a week, and over a calendar year	Кеер	Кеер

Centers continue to score higher than homes on the streamlined SAS

As expected, because we cut the easiest-to-achieve indicators in the SAS (i.e., indicators for which nearly all programs received points), we see similar patterns in the distribution of ratings comparing the original and revised streamlined SAS. In the streamlined approach, centers continued to score higher than home-based providers (see Figure 4-4). Centers continued to have a skewed distribution, with more programs rated at the higher end of the scale. In fact, 9% of all center-based programs receive the maximum score on the streamlined SAS (32 points). Homes have a more normal distribution with many programs scoring in the mid range, and fewer programs score in the low or high range.



Figure 4-4. Distribution of Total Streamlined SAS Scores by Provider Type

Note. N = 2,066 centers, 1,455 homes. Total possible points is 32 for centers and 27 for homes. *Source:* Calculations based on data provided by the Michigan Department of Education, Office of Great Start.

New alternate cuts cores for the streamlined SAS create three levels of quality for GSQ

Given the scores on the streamlined SAS, we applied three sets of cut scores to create three new levels of quality in GSQ: Level 1-Improvement Planning, Level 2-Improvement Progress, and Level 3-High Quality. Cut scores were created separately for centers and homes given that the indicators function differently for the two types of programs in GSQ. We examined three different cut score approaches: 1) the lower cut score approach; 2) the higher cut score approach; and 3) terciles. We chose these three sets of cut scores by reviewing the distribution of scores on the streamlined approach to identify any natural cut points. We also considered whether the lowest cut scores would represent too few programs or if the highest cut scores would include too many programs. We also considered the cut scores used in the original SAS to differentiate five levels of quality when considering where to place optional cut scores for only three tiers. The three sets of cuts scores are described in Table 4-11.

Scoring Approach		Level 1- Improvement Planning	Level 2- Improvement Progress	Level 3- High Quality
Lower Cut Scores	Centers	0 - <15	15 - <25	25 - 32
	Homes	0 - <9	10 - <19	19 - 27
Higher Cut Scores	Centers	0 - <19	20-<28	28 - 32
	Homes	0 - <15	15-<22	22 - 27
Terciles	Centers	0 - <27	27 - <31	31 - 32
	Home	0 - < 12	12 - <15	15 - 27
Source. Author analyses.				

Table 4-11. Three Possible Cut Scores Using the Streamlined Approach for Centers and Homes

Among the three alternate cut score approaches, the higher cut scores worked well for centers and the lower cut scores worked well for homes. Using the higher cut scores was too conservative for homes and only highlighted 132 (9%) programs as high-quality. Conversely, the lower cut score approach was too lenient for centers and highlighted over three quarters of programs as high-quality (n = 1,562, 76%). Applying the lower cut scores for homes and the higher cut scores for centers shifted the distribution of quality across the state (see Figure 4-5).

Figure 4-5. Center Ratings Using the Higher Cut Scores and Home Ratings Using the Lower Cut Scores for the Streamlined Approach



Note. N = 2,066 centers; 1,455 homes

Source: Calculations based on data provided by the Michigan Department of Education, Office of Great Start.

The streamlined approach yielded scores that were generally similar in how they grouped programs as low-, moderate-, and high-quality (see Figure 4-6). To compare the original SAS and the streamlined SAS, we grouped programs into low-, moderate-, and high-quality groups, described below:

- Low-quality
 - Scored a 1- or 2-star rating on the original SAS
 - Scored a Level 1- Improvement Planning rating on the streamlined SAS
- Moderate-quality
 - Scored a 3-star rating on the original SAS
 - Scored a Level 2- Improvement Progress rating on the streamline SAS
- High-quality
 - Scored a 4- or 5-star rating on the original SAS
 - Scored a Level 3-High Quality rating on the streamlined SAS

For centers, more programs were rated as low- or moderate- quality, while fewer programs were rated as high-quality in the streamlined approach. For homes, fewer programs were rated as low-quality in the streamlined approach, while more programs were rated as moderate- to high-quality (see Figure 4-6).

Figure 4-6 Number of Programs Rated as Low, Moderate, and High Quality in the Original SAS Versus the Streamlined Approach, for Centers and Homes



Note. N = 2,066 centers; 1,455 homes *Source.* Calculations based on data provided by the Michigan Department of Education, Office of Great Start.

The Few and Mighty approach

The second alternate scoring approach reduced the number of SAS indicators to only 7. In this approach, we selected indicators which reflect characteristics of early childhood education and care programs known from the literature to be associated with stronger program quality and child outcomes. These indicators were selected based on expert panelist feedback, where we specifically asked experts to recommend their top five items, the exploratory factor analyses and Rasch analyses presented earlier, and the review of the literature that documented the aspects of quality most closely related to child outcomes. Figure 4-7 presents the Few and Mighty indicators in the new alternate scoring approach. As noted in the figure, we included two recommendations with these revisions: (1) further developing an indicator to measure high quality, developmentally appropriate instruction, and (2) revising the consistent staff indicator to better reflect consistency of staff in programs.





More than half of all centers self-rated as high quality in the Few and Mighty approach, which was a far higher percentage than that for homes.

Working from the extant data, we selected seven specific indicators from the existing SAS that best represent the Few and Mighty indicators described in Figure 4-6. For each indicator, programs either met the criterion of this indicator or did not meet the criterion. Then, we defined those programs that met all the criteria as "high-quality" programs and those that did not meet all the criteria as "improving quality" programs.

Over half (56%) of the center-based programs achieved the "high quality" rating, while only 9% of the home-based programs achieved that rating using the Few and Mighty approach (see Figure 4-8). We also tested the importance of each of the seven Few and Mighty indicators by

sequentially removing one indicator at a time from the list and examining how each removal affected the percentages of center- and home-based programs achieving the high-quality rating. Results suggest that if we remove the "screening" indicator from the list, more centers (6% more) and homes (8% more) will achieve the high-quality rating. However, removing other indicators does not significantly affect the percentages of center- based programs or home-based programs that achieve the high-quality rating.

Figure 4-8. The Percentages of Center- and Home-Based Programs That Met and Did Not Meet All Few and Mighty Indicators for the "High Quality" Rating



For example, when we compared the published GSQ ratings with the Few and Mighty selfreport ratings, 73 percent of centers self-rated as high quality in the original SAS (those that self-rated a 4- or 5-star rating), while 56 percent would be rated high quality using the Few and Mighty approach. For homes, the original SAS highlighted 12 percent of programs as high quality (those that self-rated a 4- or 5-star rating) while 9 percent would be rated as high quality in the Few and Mighty approach.

Given these findings, the Few and Mighty indicators may be a successful strategy to highlight high quality home-based providers in Michigan. However, nearly half of the centers selfreported that they meet these key aspects of program quality. Because so many programs selfrated as high quality, the Few and Mighty approach may not give Michigan enough information to differentiate the quality of center-based care across the state. **FINDINGS SUMMARY:** Based on the previous finding that there were not five factors as expected in the SAS, the research team examined two different approaches: streamlined and the "Few and Mighty." Specifically, the "streamlined" approach reduced the number of indicators to focus on those indicators with the most variability across programs. The "Few and Mighty" approach made a more severe reduction in the total number of indicators to focus on a set of indicators with high variation and grounded in the literature describing what matters most for quality in early childhood programming. In this revised approach, the number of quality tiers reduced from five to three levels: Level 1 – Improvement Planning, where programs that met all licensing requirements are beginning the process of planning for improvement; Level 2 – Improvement Progress, for programs in the process of improvement; and Level 3 – High Quality, for programs that met high-quality standards, as verified through an observation process.

In the streamlined approach, Centers continue to score higher than homes on the streamlined SAS. For centers, more programs were rated as low- or moderate- quality, while fewer programs were rated as high-quality. For homes, fewer programs were rated as low-quality, while more programs were rated as moderate- to high-quality.

The Few and Mighty approach reduced the number of SAS indicators to only 7. These items focused on: developmental screening and use of child assessment data, culturally-relevant written plan for teaching children, family and community partnership, formal and diverse communication with families, use of and ongoing training of approved curriculum, provision of high-quality instruction, and structure and scheduling of staff to provide consistency for children. This approach found fewer centers and homes as high quality than the original SAS 5-star ratings, but many more centers than homes would be designated as high quality under this approach.

1.5 What is the variation and pattern of program-level ratings within and across program types that ensures the ratings are functioning as intended?

Over forty percent of all early childhood programs in Michigan participate in the state's voluntary QRIS. Programs with an alternate pathway in the GSQ participated at a much higher rate.

Across the state of Michigan, 41% of licensed early childhood programs have an active, published GSQ rating as of June 2018, the date we pulled all GSQ records for these analyses. Center-based programs participate at slightly higher rates (approximately 47%), compared to 40% of registered family home-based programs and 33% of licensed group child care homes. Participation does not meaningfully vary by program size; however, Great Start to Readiness Programs (GSRP), Head Start, and National Association for the Education of Young Children (NAEYC) accredited programs had higher participation rates than other program types (above 75%). These higher participation rates are most likely due to the alternate pathways afforded to these three types of programs, which allow them to skip the self-assessment process and enter GSQ with a 4-star rating. Participation is also relatively even by region within Michigan, with participation the highest in more rural regions like Northeast Michigan and the Upper Peninsula (63% and 59%, respectively) and slightly lower in the Central region and in Oakland and Macomb Counties (40% and 40%, respectively). See Tables B1 and B2 in Appendix B.

Most programs demonstrate moderate to moderate-high quality with a final published GSQ rating of 3 or 4 stars; very few programs receive the highest rating (5 stars) or the lowest ratings (1 or 2 stars).

Overall, a 3-star rating is the most common, final published rating for all programs participating in the GSQ (see Figure 4-9). Nearly one-half of all participating programs are rated as 3 stars (45%). About one-third of programs have a final rating of 4 stars (35%). Few programs receive the highest (6%) or lowest (3%) ratings in the system.



Overall, center-based programs have higher ratings than home-based programs.

Overall, center-based programs are rated higher than home-based programs (including licensed family and group child care homes) in the GSQ. Nearly 70% of all center-based programs participating in the GSQ received a final rating of 4 or 5 stars, compared to less than 10% of all home-based programs (see Figure 4-10). The most common rating for centers is 4 stars, whereas the most common rating for homes is 3 stars. This pattern in the ratings could indicate that there is an actual difference in quality between center-based and home-based programs; however, it could indicate that the GSQ is not functioning in the same way for center- and home-based programs. For example, the current measures (including the self-assessment items and/or the direct observation of quality) as currently designed may not capture unique aspects of "quality" that we would expect in these different settings.



Figure 4-10. Final Published GSQ Ratings by Program Type

The most common rating for the self-assessment survey was 3 stars, but the rating differed for centers and homes.

As described earlier, the GSQ relies on ratings from both a self-assessment survey and a direct observation of quality.⁴ Each program or classroom typically begins the GSQ rating process by completing an SAS. Programs that self-rate as 4 or 5 stars on the SAS and that want to achieve a final GSQ rating of 4 or 5 stars must register for an independent classroom observation using the PQA. This section describes each of the SAS and PQA quality measures.

Of the 3,821 programs⁵ with a final SAS rating, most programs self-rated at 3 stars (42%), and nearly one-third self-rated at 5 stars (32%). Consistent with the distribution of the final published GSQ ratings, center-based programs rated themselves higher than home-based providers (including group and family child care homes; see Figure 4-10). The most common self-rating for center-based programs was 5 stars (55%), compared to 3 stars for home-based providers (67% of group child care homes and 55% of family child care homes; see Figure 4-11).

In addition to the difference between center- and home-based care, other key program characteristics revealed consistent differences in SAS ratings. For example, programs that care for infants and toddlers had consistently lower ratings than those that do not. Accreditation status was also related to higher self-ratings, as programs with NAEYC or National Association

⁴ In general, programs complete a multistep process to achieve a final GSQ star rating, although the process varies slightly for programs that enter through an alternate pathway (e.g., Head Start, Early Head Start, and Great Start to Readiness programs).

⁵ The Michigan QRIS dataset includes 25 programs in the process of completing the GSQ process as of June 14, 2018. At the time of the analysis, these programs had a final SAS rating and were included in all SAS descriptive statistics; however, these programs had not yet received a published rating and were excluded from all final GSQ descriptive statistics. These programs were eligible for the validation study sample.

for Family Child Care (NAFCC) accreditation consistently had higher self-ratings than nonaccredited programs.





FINDINGS SUMMARY: Over 40% of all early childhood programs in Michigan participate in GSQ. Programs with an alternate pathway in the GSQ participated at a much higher rate (about 75% participation rate). Most programs demonstrate moderate to high quality with a final published GSQ rating of 3 or 4 stars; very few programs receive the highest rating (5 stars) or the lowest ratings (1 or 2 stars). Center-based programs are rated higher than home-based programs, with nearly 70% of all center-based programs receiving a final rating of 4 or 5 stars, compared to less than 10% of all home-based programs.

2.1 Are there differences in GSQ ratings based on other measures of quality using the Environment Rating Scales (ERS) and Classroom Assessment Scoring System (CLASS) across program types?

To assess how GSQ ratings align with other measures of quality, we compared the scores from independent observations of quality at each GSQ star level, using analysis of variance (ANOVA). Our analyses compare GSQ star ratings with a program's respective scores on the ERS and CLASS. We expected to see positive relationships between the GSQ ratings and the ERS and CLASS, with 4- and 5-star programs rating higher than 1-, 2-, and 3-star programs.

GSQ programs are rated as providing minimal to good quality care, with high ratings in emotional support and interactions.

Descriptive analyses were conducted for the measures of quality, ERS and CLASS. For the ERS, a score of 1 represents Inadequate care, a score of 3 represents Minimal care, a score of 5 represents Good care, and a score of 7 represents Excellent care. Analyses indicated that programs were performing above the Minimal care level on the ERS regardless of program type or age group (see figure 4-12). In particular, programs performed at or above the 4 range on Interactions, Language and Literacy (only for home-based programs), and Program Structure (programs serving preschool-age children), which is in between minimal and good. There was a general trend that center-based programs serving preschool-age children performed better on the ERS than home-based programs and center-based programs serving infants and toddlers, with one exception. (Home-based programs performed better than other programs in the Learning Activities subscales.) It is important to note that programs generally scored the lowest in Learning Activities, below the minimal level. Caution should be taken in interpreting the findings with infants in light of their small sample size.



Figure 4-12. Environment Ratings Scale Ratings Across Type of Tool

Note. N = 182. ERS = Environment Rating Scale. *Source.* Validation Study Data as of December 31, 2018 Turning to the CLASS measure, which was only conducted in center-based preschool classrooms, we see that programs generally performed well in Emotional Support and Classroom Organization, with an average of 5.6 and 5.0, respectively, out of a 1–7 scale. Emotional Support is based on the dimensions of positive climate, negative climate, teacher sensitivity, and regard for student perspectives. Classroom Organization is based on the dimensions of behavior management, productivity, and instructional learning formats. On average, programs scored 2.86 in Instructional Support on a scale ranging from 1–7, indicating low quality of instruction and feedback. Instructional Support is based on the dimensions of concept development, quality of feedback, and language modeling. It is important to note that programs tend to score considerably lower on the Instructional Support domain compared to the Classroom Organization and Emotional Support domains. Researchers have indicated that the Instructional Support domain is most related to children's outcomes, with stronger findings at the threshold of 3.25 or higher (e.g., Burchinal, Vandergrift, Pianta, & Mashburn, 2010). Forty percent of programs had a score of 3.25 or higher on Instructional Support.



Note. N = 65. CLASS = Classroom Assessment Scoring System.

Source. Validation Study Data as of December 31, 2018

What is the relationship between star rating and ERS and CLASS Scores?

Results indicate a moderate relationship between the GSQ star rating and ERS, and small relationship between GSQ star rating and CLASS scores.

We first examined the bivariate correlation between the GSQ star rating and ERS scores across program type (center, group child care home, family child care home; see Table 4-12). We see moderate correlations between ERS total and rating across program type, with the rating being stronger for home-based programs; this pattern generally maintains itself for the ERS subscales. The correlation between the CLASS domains and GSQ star rating ranged from .16 (Classroom Organization) to .17 (Emotional Support and Instructional Support), indicating minimal relationship between GSQ rating and CLASS domain.

		Ratings	
Variables	LCCC	LGCCH	RFCCH
ERS total	0.267	0.517	0.384
Space and furnishing	0.290	0.320	0.394
Personal care routines	0.127	0.389	0.309
Rating Language and literacy	0.221	0.364	0.362
Learning activities	0.292	0.451	0.381
Interaction	0.216	0.444	0.251
Program structure	0.201	0.590	0.270

Table 4-12. Correlation ERS and Ratings

Note. CLASS only collected for pre-K classrooms. ERS = Environmental Rating Scale.

LCC = Licensed Child Care Center; LGCCH = Licensed Group Child Care Home; RFCCH = Registered Family Child Care Home

Source. Validation Study Data as of December 31, 2018

Are there significant differences in independent observations of quality by GSQ ratings?

Results find that GSQ star ratings are related to some, but not all, scores on the independent observations among centers and home-based programs.

Examination of the ERS total score across star levels indicates a general pattern where 4- and 5star programs score higher than 1-, 2-, and 3-star programs (see Figure 4-14). To empirically examine whether there were significant differences in the independent observations of quality by GSQ ratings, we collapsed the star levels and program types. This was done due to smaller than expected sample size. The 1-, 2-, and 3-star programs were collapsed into a "low star" group and the 4- and 5-star programs were collapsed into the "high star" group; the group and family home programs were collapsed into a "home-based" program group. Analysis of variance (ANOVA) was conducted to examine whether there were significant differences by star rating and program type for the ERS scores. Results indicated that there were significant differences by star ratings for the ERS total score and subscales. There were also differences by program type (center vs. home-based programs; see Table 4-13). Centers scored higher than group child care homes in the area of Personal Care Routines (there was a trend at p <.10 that centers also scored higher on Space and Furnishing). However, family child care homes scored higher than centers in the areas of Language and Literacy and Learning Activities.





Table 4-13. Summary of ANOVA Analyses Examining Differences in ERS by Program Type andStar Rating

	Program Type Differences	Star Rating Differences		
ERS Total	No	Yes^		
Space and Furnishing	No	Yes^		
Personal Care Routines	Yes*	Yes^		
Language and Literacy	Yes**	Yes^		
Learning Activities	Yes**	Yes^		
Interaction	No	Yes^		
Program Structure	No	Yes^		
Note. N = 182. ERS = Environment Rating Scale.				

^ = favors 4- and 5-star programs

* = favors center-based programs

** = favors home-based programs

Note. N = 182 *Source.* Validation Study Data as of December 31, 2018

Similarly, there was a trend with 5-star programs scoring higher on the CLASS than 4-star programs, and 4-star programs scoring higher than 3-star programs (see Figure 4-15). Caution should be taken in interpreting the data for 2-star center-based programs due to their small sample size. (*Note:* There are no 1-star center-based programs). Due to the small sample size at each star level, inferential statistics were not conducted. Instead, we collapsed center- based programs into low (1-, 2-, and 3-star programs) and high (4- and 5-star programs). These inferential analyses indicated no significant difference across low- and high-quality programs; however, there was a trend (p < .10) that 4- and 5-star programs scored higher than 2- to 3-star programs on Emotional Support and Classroom Organization. Due to the sample size, caution should be taken in interpreting these findings.



Figure 4-15. Average CLASS Domain Scores by Star Rating

Note. N = 65 *Source.* Validation Study Data as of December 31, 2018

FINDINGS SUMMARY: Two classroom observation measures – the Environment Rating Scale (ERS) and Classroom Assessment Scoring System (CLASS) — were used to independently evaluate the GSQ ratings. On average, GSQ programs were rated as providing minimal to good quality care with the ERS. Preschool classrooms also assessed with the CLASS were rated as moderate to high quality on the Emotional Support and Classroom Organization of the CLASS and low on the Instructional Support. Correlation analyses found a moderate relationship between the GSQ star rating and ERS, and small relationship between GSQ star rating and CLASS scores. Results indicated that there were significant differences by star ratings for the ERS total score and subscales with 4- and 5-star programs being rated higher than 1-, 2-, and 3-star programs. A similar trend was found for preschool programs based on the CLASS, but it was not significant. Caution should be taken due to the small sample size of preschool classrooms.

2.2 How well do the ratings from the alternate cut points and rules distinguish different levels of quality?

Next, we examined the relationship between the two alternate scoring approaches (streamlined and Few and Mighty) and the CLASS and ERS data collected for the validation study to answer *RQ 2.2: How well do the ratings from the alternate cut points and rules distinguish different levels of quality?* To address this question, we examined the extent to which simulated SAS ratings (developed for RQ 1.4) differed across the three levels of rating. The analyses mirrored those described for RQ 2.1 but used the simulated alternate scoring approaches. For these analyses we compared the mean CLASS and ERS scores for programs rated as high quality on the simulated scores with programs that rated lower (i.e., Level 1-Improvement Planning or Level 2-Improvement Progress). These analyses used CLASS and ERS data collected as of December 31, 2018 and included 176 total programs (65 with CLASS and ECERS, 44 with ITERS, and 104 with FCCERS).

The streamlined approach consistently distinguishes high-quality programs from moderateor low-quality programs and works best for home-based providers.

Overall, there is some support that the streamlined approach identifies meaningful differences in quality as measured by the CLASS and ERS observations. Generally, programs rated as Level 3-High Quality scored higher on the CLASS and ERS than programs that were rated as Level 1-Improvement Planning or Level 2-Improvement Progress (see Table 4-14). However, there were some differences by centers and homes and by the age of children served.

- For centers serving preschool-age children, programs rated as Level 3-High Quality scored higher than those at the lower levels on the streamlined approach on both the CLASS and the ECERS. However, there were no statistically significant differences in quality ratings between Level 1-Improvement Planning or Level 2-Improvement Progress. This suggests that the streamlined approach accurately identifies the highest quality center-based programs, but does not differentiate low from moderate quality.
- For home-based providers, the streamlined approach differentiated all levels of quality. Specifically, programs rated as a Level 3-High Quality scored significantly higher than programs rated as a Level 2-Improvement Programs and programs rated as a Level 2-Improvement Progress scored significantly higher than programs rated as a Level 1. This suggests that the streamlined approach accurately identifies all levels of quality for home-based programs.
- However, for centers that serve infants and toddlers, there were no statistically significant differences in ITERS scores among centers at different levels of quality, suggesting it may not work well for programs serving our youngest learners.

These findings suggest that the streamlined approach works best for home-based programs serving older children. Interpretations should be made cautiously given the small sample sizes for the center-based comparisons, which sometimes include as few as 18 programs.

	Scoring Approach	Level 1- Improvement Planning	Level 2- Improvement Progress	Level 3- High Quality_
CLASS		N = 11	N = 29	N = 25
	Emotional Support	5.39 (1.07)	5.46 (1.06)	5.90 (0.94)
	Classroom Organization	4.76 (0.94)	4.74 (1.31)	5.31 (1.13)
	Instructional Support	2.76 (1.29)	2.52 (1.16)	3.31 (1.43)
	Total score ^a	4.30 (1.01)	4.24 (1.07)	4.84 (1.05)
ERS (Centers)				
ECERS		N = 11	N = 30	N = 24
	Space & Furnishing	3.54 (0.61)	3.52 (1.00)	4.07 (0.77)
	Personal Care Routine	3.41 (0.97)	3.47 (1.24)	4.05 (1.07)
	Language & Literacy	3.58 (1.00)	3.74 (1.23)	4.12 (0.98)
	Learning Activities	2.03 (0.48)	2.43 (0.89)	2.64 (0.74)
	Interaction	4.50 (1.26)	4.45 (1.56)	5.16 (1.12)
	Program Structure	3.80 (1.61)	3.85(1.67)	4.59 (1.07)
	Total Score ^a	3.49 (0.76)	3.58 (1.09)	4.09 (0.72)
ITERS		N = 8	N = 19	N = 17
	Space & Furnishing	3.05 (1.16)	3.25 (0.95)	3.44 (1.42)
	Personal Care Routine	3.23 (1.15)	3.06 (0.79)	3.42 (1.13)
	Language & Literacy	3.65 (1.32)	3.71 (1.33)	4.00 (0.94)
	Learning Activities	1.99 (0.70)	2.19 (0.66)	2.21 (0.57)
	Interaction	3.82 (1.39)	4.36 (1.59)	4.74 (1.14)
	Program Structure	3.64 (1.52)	3.86 (1.96)	4.05 (1.77)
	Total Score ^c	3.23 (1.12)	3.41 (0.97)	3.64 (0.93)
ERS (Homes)				
FCCRS		N = 30	N = 47 - 48	N = 27
	Space & Furnishing	2.59 (0.89)	3.07 (0.92)	3.92 (1.28)
	Personal Care Routine	2.32 (0.86)	2.89 (1.11)	3.29 (1.04)
	Listening & Talking	3.49 (1.41)	3.83 (1.48)	5.21 (1.16)
	Learning Activities	2.14 (0.83)	2.64 (0.92)	3.30 (1.11)
	Interaction	4.02 (1.81)	4.61 (1.66)	5.69 (1.38)
	Program Structure	2.74 (1.56)	3.68 (1.99)	4.65 (1.66)
	Total Score ^b	2.88 (1.02)	3.42 (1.08)	4.29 (0.93)

Table 4-14. CLASS and ERS Scores by Level Rating in the New Streamlined Approach

Note. CLASS only collected for PreK classrooms. ^a Means at Level 3 are statistically different from means at Level 2 and Level 1, but the means at Level 2 are not statistically different from the means at Level 1. ^b Means at all three levels are statistically different from one another (i.e., the mean at Level 3 is statistically greater than the mean at Level 2 which is statistically greater than the mean at Level 1). ^c No statistically significant findings were detected. *Source.* Validation Study data as of December 31, 2018

The ratings from the Few and Mighty approach do well to distinguish two distinct levels of quality for home-based programs, but do not for center-based programs.

Overall, there is support that the Few and Mighty approach is picking up on meaningful differences in quality for home-based programs as measured by the ERS, but not for centers (see Table 4-15).

- Home-based programs rated as Level 3-High Quality using the Few and Mighty approach have statistically significant and higher ERS scores than those rated as Level 2-Improving Quality.
- For centers, the average scores were in the hypothesized direction with higher CLASS and ERS (including ECERS and ITERS) scores for those rated as Level 3-High Quality, but these differences were not statistically significant. This may be explained in part by the small sample sizes for the center-based comparisons, which sometimes include as few as 18 programs.

These findings indicate that the Few and Mighty scoring approach does well at distinguishing two levels of quality for home-based programs, but not for center-based programs.

FINDINGS SUMMARY: The streamlined and Few and Mighty approaches were empirically examined using the ERS and CLASS. The streamlined approach consistently identified highquality programs from moderate- or low-quality programs and worked best for home-based providers. Generally, programs rated as Level 3-High Quality scored higher on the CLASS and ERS than programs that were rated as Level 1-Improvement Planning or Level 2- Improvement Progress. The streamlined version differentiated between all levels of high quality for homebased programs, between Level 3-High Quality and Level 2-Improvement Progress, and Level 2-Improvement Progress and Level 1-Improvement Planning.

The Few and Mighty approach distinguished between two distinct levels of quality for homebased programs, but not for center-based programs. Specifically, home-based programs rated as Level 3-High Quality have higher ERS scores than those rated as Level 2-Improving Quality. Centers in the Level 3-High Quality group had higher ERS and CLASS scores than in the other groups, but these differences were not statistically significant, which may be due to the small sample size.

	Scoring Approach	Improving Quality	High Quality
CLASS		N = 39	N = 26
	Emotional Support	5.54 (1.06)	5.74 (0.98)
	Classroom Organization	4.84 (1.20)	5.14 (1.20)
	Instructional Support	2.71 (1.27)	3.09 (1.38)
ERS (Centers)		N = 39	N = 26
ECERS	Space & Furnishing	3.63 (0.84)	3.88 (0.96)
	Personal Care Routine	3.61 (1.05)	3.78 (1.32)
	Language & Literacy	3.73 (1.14)	4.04 (1.05)
	Learning Activities	2.33 (0.80)	2.62 (0.77)
	Interaction	4.55 (1.36)	4.99 (1.40)
	Program Structure	3.89 (1.56)	4.48 (1.31)
	Total Score	3.61 (0.92)	3.96 (0.95)
ITERS		N = 26	N = 18
	Space & Furnishing	3.17 (1.00)	3.46 (1.40)
	Personal Care Routine	3.08 (0.86)	3.44 (1.15)
	Language & Literacy	3.61 (1.30)	4.11 (0.93)
	Learning Activities	2.15 (0.68)	2.17 (0.55)
	Interaction	4.19 (1.44)	4.73 (1.32)
	Program Structure	3.68 (1.77)	4.20 (1.80)
	Total Score	3.31 (0.99)	3.68 (0.93)
ERS (Homes)		N = 81	N = 23
FCCERS	Space & Furnishing	2.96 (1.02)	3.88 (1.21)
	Personal Care Routine	2.72 (1.06)	3.24 (1.07)
	Listening & Talking	3.82 (1.49)	5.10 (1.24)
	Activities	2.45 (0.93)	3.44 (1.04)
	Interaction	4.44 (1.72)	5.71 (1.45)
	Program Structure	3.25 (1.87)	5.08 (1.34)
	Total Score*	3.24 (1.09)	4.37 (0.91)

Table 4-15. CLASS and ERS Scores by Level Rating in the New Few and Mighty Approach

Note. * Indicates statistical significance as the p < 0.05 level

CLASS = Classroom Assessment Scoring System; ERS = Environment Rating Scale; ECERS = Early Childhood Environment Rating Scale; ITERS = Infant-Toddler Environment Rating Scale; FCCERS = Family Child Care Environment Rating Scale

Source. Validation Study data as of December 31, 2018

V. Discussion

The goal of this validation study was to examine (1) if and how GSQ measures "quality" in licensed centers and for home-based providers; and 2) the difference between ratings in terms of quality. Through a multi-method, multi-informant process that involved analyses of extant data, independent observations, surveys, and reliability checks, we were able to examine these questions. In particular, we sought to address seven specific questions. These questions include

- 1. Are the QRIS quality components and standards the "right" ones?
- 2. Is the process of documenting and verifying each indicator yielding reliable and accurate results?
- 3. What is the relationship among the components and are they functioning as expected?
- 4. What are the most appropriate ways to combine measures of quality standards into summary ratings when examining the cut scores and combining rules?
- 5. What is the variation and pattern of program-level ratings within and across program types that ensures the ratings are functioning as intended?
- 6. Are there differences in GSQ ratings based on other measures of quality using the Environment Rating Scales (ERS) and Classroom Assessment Scoring System (CLASS) across program types?
- 7. How well do the ratings from the alternate cut points and rules distinguish different levels of quality?

A comprehensive review of the questions and findings indicate the following:

- Engagement with experts in the field did not provide clarity as to the specific indicators of the SAS that should be kept to improve quality and child outcomes. Though experts agreed that many of the components in the SAS are strongly linked to program quality rather than child outcomes, they did not agree on which SAS indicators. This discordance between experts may be due to their diverse perspectives, which included expertise in QRIS evaluation, professional development, infant/toddler practices, home-based practices, and birth-to-five administration.
- The process of documenting and verifying each indicator is reliable and sound, indicating the standards are clear. Specifically, an independent process of reviewing the SAS and accompanying documentation resulted in similar points and ratings for programs. This indicates a clear alignment between standards and the process of verification.
- The SAS is moderately associated with overall star rating, with over two-thirds of programs' self-rating highly correlated with their official rating. This means that the process of programs completing the self-rating is reliable for a majority of programs, but there are many instances where programs overrate themselves (rate themselves higher than they should be rated).

- The SAS, which includes five areas thought to be critical for examining quality linked to child outcomes, contains many "easy" items that contribute to the psychometric weakness of the survey. Exploratory factor analyses did not confirm the five factors, nor did the Rasch analyses. The Rasch analyses further confirmed that there were too many "easy" items and suggested the need for a variety of easy, moderate, and difficult level indicators.
- Even with the concern about the factor structure of the SAS, there is an indication that there are meaningful differences in GSQ ratings based on independent measures of quality using the ERS and CLASS. There was a general trend that higher rated programs had higher average scores on the ERS and CLASS compared to lower rated programs. There were few differences by program type on the ERS, with some of the scores favoring family child care homes. There were no significant differences by star rating on CLASS, but results indicated a trend of higher rated programs performing better than lower rated programs on Emotional Support and Classroom Organization, but no evidence of differences in Instructional Support. However, some caution in interpreting this data is advised due to the small sample size, especially for 2-star centers and 5-star programs.

a. Limitations

Caution should be taken in generalizing these findings due to the sample sizes in certain strata (e.g., 1-, 2-, and 5-star centers). While we have tried to account for some of this imbalance by combining strata (e.g., 1-, 2-, and 3-star programs), this method limits our ability to be precise about meaningful differences in star levels. While the ERS and CLASS are established measures that include training developed by the creators of the instruments, there is still potential for rater error and drift, though the issue of drift is somewhat minimal due to the short window of this data collection. Considering the low response rate for this study, caution should be taken before concluding that those who volunteered to participate in this study are similar to those who actively or passively declined. There is a need to gather more data from programs serving infants/toddlers across the star rating level to understand the impact of those programs on star ratings. Finally, this validation study did not include child outcomes; therefore, we are unable to confirm the extent to which GSQ star rating levels are associated with child outcomes and growth in children's learning and development.

VI. Recommendations

In general, the results from the GSQ Validation Study suggest that the current GSQ system is functioning largely as intended. Analyses indicated that the GSQ star rating level is positively and significantly associated with higher levels of observed classroom quality, albeit small to moderate. Additionally, a series of ANOVAs found significant differences by star rating level for some aspects of observed quality; however, the lack of child outcomes limits the veracity of the claims about the validity of GSQ.

An in-depth analysis of extant data indicated that adjustments to the SAS and how it is placed in the system could strengthen the effectiveness of the GSQ in promoting and maintaining program quality and child outcomes. Recommendations for modifications to the system are based on a system-wide approach and incorporating what experts are saying are the "Few and Mighty" measures that can support program quality and child outcomes. We sought to create an SAS that is psychometrically sound through validation with observed measures of quality.

We recommend the following:

- Determine how to collect child outcomes data to further bolster changes to the system. While there is no robust evidence about the link between QRIS and child outcomes, with some QRIS finding more effects than others, there is a need for some child outcomes data as another marker of validation of the GSQ. Consideration should be given as to how this can be done prior to implementing the next iteration of the system.
- Determine if programs are consistently interpreting and providing evidence to meet the SAS indicators at levels we would expect. For future SAS revisions, we recommend that MDE build in time for defining expectations, piloting, and finalizing each indicator. This process could begin by reviewing the literature and bringing in experts on links between structural and process quality and known child outcomes, defining expectations for the level of evidence programs would need to demonstrate to meet the indicator, and piloting the items with representatives from multiple program types to determine if programs are consistently interpreting and documenting program quality as expected.
- There is a need to streamline the SAS to make it more psychometrically sound and reliable. As currently crafted, the SAS does not empirically generate five factors and there is a lack of "difficult" indicators, reducing its utility. Thus, there is a need to examine the goal and role of the SAS and how it can be recrafted to better meet the needs of the system, with less of a focus on rating and more focus on continuous quality improvement. Our analyses indicated that a "Few and Mighty" approach, reducing the SAS to 7 indicators, allowed us to differentiate high-quality programs from all other programs. These items could stand on their own or serve as a foundation for future SAS revisions.
- Further examine how the SAS and GSQ are working for center-based and home-based providers. Center-based programs typically receive higher SAS ratings than home-based

programs, which allows them greater opportunities to receive higher ratings on the GSQ. Meanwhile center-based programs received slightly higher scores on average than home-based programs on the classroom observations, especially in the Personal Care Routines Subscale of the ERS; however, home-based programs are rated higher in Language and Literacy and Learning Activities subscales of the ERS. Thus, MDE may want to consider how to equitably weight various components of the SAS to ensure fairness across program types.

• Further examine how programs serving infants/toddlers are viewing the system in comparison to those serving preschool-age children. As star ratings are assigned at a program level, there is a need to examine the extent to which certain age grouping may hinder or enhance a program's star rating. In this study, there is indication that programs serving infants and toddlers may score a little lower, potentially harming the availability of classrooms serving very young children.

VII. References

- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental measurement in the human sciences, Second Edition.* Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Burchinal, M. (2008). How measurement error affects the interpretation and understanding of effect sizes. *Child Development Perspectives, 2,*178–188.
- Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre- kindergarten programs. *Early Childhood Research Quarterly, 25*(2), 166–176. doi:10.1016/j.ecresq.2009.10.004
- Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., . . . Barbarin, O. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood Research Quarterly*, 21(2), 174-195. doi:10.1016/j.ecresq.2006.04.004
- Early, D. M., Sideris, J., Neitzel, J., LaForett, D. R., & Nehler, C. G. (2018). Factor structure and validity of the Early Childhood Environment Rating Scale – Third Edition (ECERS-3). *Early Childhood Research Quarterly*, 44, 242–256. doi:https://doi.org/10.1016/j.ecresq.2018.04.009
- Harms, T., & Clifford, R. (2007). *Family Child Care Environment Rating Scale Revised.* New York, NY: Teachers College.
- Harms, T., Clifford, R., & Cryer, D. (2015). *Early Childhood Environment Rating Scale(3rd ed.).* New York, NY: Teachers College.
- Harms, T., Cryer, D., Clifford R., & Yazejan, N. (2017). *Infant/Toddler Environment Rating Scale, Third Edition (ITERS-3).* New York, NY: Teachers College.
- HighScope Educational Research Foundation. (2003). *Preschool program quality assessment,* (*PQA*) administration manual (2nd ed.). Ypsilanti, MI: HighScope.
- Michigan Department of Education (2007). A great start for kids: Recommendations for a Michigan child care Quality Rating and Improvement System. Retrieved from http://greatstartforkids.org/sites/default/files/file/QRS_FinalB_2007.pdf
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide. Eighth Edition.* Los Angeles, CA: Muthén & Muthén.

Tout, K., Magnuson, K. Lipscomb, S., Karoly, L., Starr, R., Quick H., ... & Wenner, J. (2017).
 Validation of the quality ratings used in Quality Rating and Improvement Systems (QRIS):
 A synthesis of state studies. OPRE Report #2017-92. Washington, DC: Office of Planning,
 Research and Evaluation, Administration for Children and Families, U.S. Department of
 Health and Human Services.

Wright, B. D., & Masters, G. (1982). Rating scale analysis. Chicago, IL: Mesa.

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. VIII. Appendix

Appendix A. Analysis Methods and Detailed Results for Analyses Using Extant Data

This appendix includes the additional details about the analysis methods and results used to examine the relationships among the quality component measures to assesses whether they are functioning as expected. To identify the latent constructs measured by the Self-Assessment Survey (SAS), the research team conducted both an exploratory factor analysis (EFA) and a Rasch analysis.

Some survey questions appeared to be "easy items."

We examined classroom-level SAS data, rather than program-level data, in terms of survey response pattern. Two types of classrooms were examined separately and in combination: center-based classrooms (N = 2,956) and home-based classrooms (N = 1,441). Items for these two types of classrooms were slightly different: compared to the survey for home-based classrooms, the survey for center-based classrooms contained a few more center-specific questions on staff qualifications, administration, family collaboration, and a question about whether each child had a consistent team of providers/educators and peers over a week. Each item was coded into a 1 or 0 according to the design: Classrooms received one point for checking the item and received zero points for not checking the item. Table A1 presents the domains, subdomains, and items between center-based classrooms and home-based classrooms. In addition, the table includes the percentage of respondents who marked "Yes" on a given item. As shown in Table A1, responses to some items were highly skewed with response rates over 95% (for example, the family collaboration item about staff engaging in informal communication with parents, the administration item about having written personnel policies and procedures, and the physical environment item about being free of environmental risks). In addition, center-based classrooms reported more "easy items" than home-based classrooms. In contrast, most home- based classrooms found some of the items in the Staff domain and the Screening domain to be "difficult" (with a 12% to 13% "Yes" rate). Given the presence of items with few or no variations, the statistical algorithm could not correctly compute standard errors or covariate matrices for the sample; therefore, the overall results may not be trustworthy. Interpretations of results in this section should be made with caution.
Table A1. Domains, Items, and Response Rates for Center-Based Classrooms and Home-Based Classrooms

Domain Subdomain Variable Reserd Reserd Center Her Combined Staff Qualifications DirecTOR2 V V 20.0 13.2 17.7 Development Training DirecTOR3 V V 20.0 13.2 17.7 DirecTOR3 V V 40.4 15.2 12.1 38.7 Development Staff Lead STAFF1 V V 15.6 10.8 14.1 Qualifications STAFF2 V V 15.5 11.0 15.5 STAFF3 V V 15.3 14.1 15.5 14.1 Qualifications STAFF3 V V 15.3 14.1 15.5 STAFF3 V V 15.3 14.1 15.5 14.1 Qualifications STAFF9 V 12.7 na na STAFF1 V 20.2 na na 16.5 Development PD1				Center-	Home-	"Yes"	Response (Perc	entage)
Staff Qualifications and Professional DevelopmentDirector R Education EducationsDirector R 2VV11.182.034.3DevelopmentEducation Family QualificationsEducation STAFF1VV20.013.217.7DIRECTOR3VV40.415.232.1Director R 36.736.736.7Provider QualificationsSTAFF1VV15.510.814.115.5Staff Assistant QualificationsSTAFF3VV15.510.415.515.7Staff Assistant QualificationsSTAFF3VV13.311.1015.515.7Staff Assistant QualificationsSTAFF3VV12.3na </th <th>Domain</th> <th>Subdomain</th> <th>Variable</th> <th>Based</th> <th>Based</th> <th>Center</th> <th>Home</th> <th>Combined</th>	Domain	Subdomain	Variable	Based	Based	Center	Home	Combined
and Professional DevelopmentDIRCTOR3vvv0.01.21.7.DevelopmentTraining DIRECTOR1VV40.41.5232.1Staff LeadSTAFF1VV1.2.980.435.1ProviderSTAFF2VV1.5.51.0.81.1.1ProviderSTAFF3VV1.5.51.0.81.1.1ProviderSTAFF3VV1.5.91.4.71.5.5Staff AssistantSTAFF6V1.2.3n.an.aQualificationSTAFF9V2.0.3n.an.aStaff AssistantSTAFF9V2.0.2n.an.aQualificationSTAFF9V2.0.2n.an.aProfessionalSTAFF9V2.0.2n.an.aProfessionalSTAFF9V2.0.2n.an.aProfessionalSTAFF9V2.0.2n.an.aProfessionalPD1VV7.2.77.2.97.4.8DevelopmentPD1VV7.5.77.2.97.4.8CommunityPD2VV7.5.77.2.97.4.8PattershipsFAMILY1VV9.8.99.9.09.9.5FAMILY2VV9.8.18.8.99.9.09.9.5FAMILY2VV9.8.18.8.39.4.9CommunityPattershipsFAMILY2VV9.8.18.8.3 <td>Staff Qualifications</td> <td>Director</td> <td>DIRECTOR1</td> <td>٧</td> <td>٧</td> <td>11.1</td> <td>82.0</td> <td>34.3</td>	Staff Qualifications	Director	DIRECTOR1	٧	٧	11.1	82.0	34.3
Development Training DIRECTORA V V 40.4 15.2 32.1 Fart Lead STAFF2 V V 12.9 80.4 35.1 Provider STAFF2 V V 15.9 14.1 15.5 STAFF3 V V 15.9 14.7 15.5 STAFF4 V V 15.9 14.7 15.5 STAFF5 V 14.6 na na na Staff Assistant STAFF3 V 20.3 na na Qualifications STAFF10 V 20.2 na na Development PD2 V V 72.9 66.9 69.7 Development PD3 V V 72.1 30.7 Staf.6 Community PD4 V V 61.9 69.7 Development FAMILY1 V V 98.9 99.0 92.5 Family and Family TV	and Professional	Education and	DIRECTOR2	v	v	20.0	13.2	17.7
International staff and and staff and and staff and s	Development	Training	DIRECTOR3	V	v	40.4	15.2	32.1
Staff Lead STAFF2 v v 12.9 80.4 33.1 Provide STAFF2 v v 15.9 14.7 15.5 Gualifications STAFF3 v v 19.3 11.0 16.5 STAFF5 v 14.6 na na na na Staff Assistant STAFF5 v 20.3 na na na Qualifications STAFF8 V 20.3 na na na Qualifications STAFF7 V 20.2 na na na Professional PD1 V V 72.9 69.7 72.9 69.7 Professional PD2 V V 72.1 30.7 58.6 PD2 V V 10.27.0 78.8 69.0 99.6 Partnerships FAMILY1 V V 98.1 88.9 95.4 FAMILY2 V V 98.1 88			DIRECTOR4	V	v	52.4	10.7	38.7
Provider Qualifications STAFF3 v v 15.6 10.8 14.1 Qualifications STAFF3 v v 19.3 11.0 15.5 STAFF6 v 14.6 ma ma ma STAFF6 v 14.6 ma ma ma Qualifications STAFF6 v 22.4 ma ma Qualifications STAFF9 v 20.3 ma ma Qualifications STAFF9 v 20.2 ma ma Patersonal PD1 v v 72.9 74.8 Po2 v v 72.1 30.7 58.6 Community PD3 v v 72.1 30.7 58.6 Patnerships FAMILY1 v v 98.8 99.0 99.6 Patnerships FAMILY2 v v 98.1 89.9 95.4 Patnerships Community V v		Staff Lead	STAFF1	v	٧	12.9	80.4	35.1
Administration and Management Qualifications STAFF3 v v 15.9 11.0 16.5 STAFF4 v v 13.3 11.0 16.5 Staff Assistant Qualifications STAFF6 v 52.4 na na STAFF8 v 20.3 na na na STAFF10 v 20.3 na na na STAFF10 v 20.2 na na na PD1 v v 72.9 62.9 69.7 PD2 v v 75.7 72.9 74.8 PD3 v v 75.7 72.9 74.8 PD4 v v 92.9 92.5 62.9 69.7 Patherships FAMILY1 v v 98.8 99.0 99.6 FAMILY2 v v 98.2 90.7 95.7 Patherships FAMILY3 v v 98.1 na		Provider	STAFF2	v	٧	15.6	10.8	14.1
Administration STAFF3 v v 19.3 11.0 16.5 STAFF5 v 14.6 ma ma ma STAFF5 v 32.3 ma ma ma Qualifications STAFF3 v 32.3 ma ma STAFF9 v 22.3 ma ma ma Professional PD1 v v 72.9 62.9 69.7 Development PD2 v v 75.7 72.9 74.8 Community PD3 v v 75.7 72.9 74.8 PD4 v v 64.0 77.8 78.6 77.8 Community Patnerships FAMILV2 v v 99.8 99.0 99.6 Patnerships FAMILV3 v v 98.1 88.9 95.4 FAMILV5 v v 96.5 ma ma ma Management <td< td=""><td></td><td>Qualifications</td><td>STAFF3</td><td>v</td><td>V</td><td>15.9</td><td>14.7</td><td>15.5</td></td<>		Qualifications	STAFF3	v	V	15.9	14.7	15.5
STAFF5 V 14.6 na na Staff Asistant STAFF6 V 52.4 na na Staff Asistant STAFF0 V 32.3 na na Qualifications STAFF1 V 20.3 na na Partice STAFF10 V 20.2 na na Porfessional PD1 V V 22.9 66.9 Development PD2 V V 75.7 72.9 67.9 Community PD4 V V 06.0 27.0 49.8 FAMILY3 V V 92.4 92.9 92.5 62.9 Community FAMILY1 V V 92.8 99.0 99.6 FAMILY3 V V 98.1 89.9 95.4 FAMILY3 V V 98.1 89.9 95.7 FAMILY6 V V 98.1 88.3 94.9			STAFF4	v	V	19.3	11.0	16.5
Image: start of the system STAFF6 V 52.4 na na Start of the system STAFF7 V 32.3 na na STAFF8 V 20.3 na na na STAFF9 V 12.7 na na na Professional PD1 V V 72.9 74.8 Porfessional PD2 V V 75.7 72.9 74.8 Post V V 75.7 72.9 74.8 74.8 Community PD2 V V 72.1 30.7 58.6 Post V V 92.4 92.9 92.5 78.7 Family and Family V V V 98.0 94.0 94.6 77.8 Community FAMILY3 V V 98.1 88.3 94.9 Community COMMUNITY3 V V 98.2 98.7 77.5 90.4 M			STAFF5	v		14.6	na	na
Staff Assistant STAFF v 32.3 na na Qualifications STAFF8 v 20.3 na na STAFF9 v 12.7 na na Professional PP01 v V 20.2 na na Development PD1 v V 72.9 66.2 66.7 Tomily and Family PD2 v v 72.9 72.9 74.8 Community Pold v v 72.1 30.7 58.6 Patherships FAMILV1 v v 92.9 92.5 92.5 FAMILV3 v v 94.0 44.6 77.8 93.6 FAMILV5 v 83.9 na na na na FAMILV5 v 98.3 94.0 95.7 77.5 90.4 Patherships FAMILV5 v v 98.7 77.5 90.4 Administratio			STAFF6	V		52.4	na	na
Qualifications STAFF8 v 20.3 na na Professional PD1 v 12.7 na na Development PD1 v v 72.9 62.9 69.7 Development PD2 v v 75.7 72.9 74.8 Porfessional PD1 v v 72.1 30.7 58.6 Po1 v v 62.9 92.5 74.8 74.8 Community Partnerships FAMILV1 v v 94.0 44.6 77.8 Partnerships FAMILV3 v v 96.5 na na FAMILV5 v 83.9 na na na na FAMILV5 v 96.5 na na na na FAMILV3 v v 96.7 77.5 90.4 38.3 Management Administration Administration Administruition na na <td></td> <td>Staff Assistant</td> <td>STAFF7</td> <td>v</td> <td></td> <td>32.3</td> <td>na</td> <td>na</td>		Staff Assistant	STAFF7	v		32.3	na	na
Administration and Management Family and Family and Community Partnerships Family Partnerships FAMILV1 V V 72.9 62.9 69.7 Panily and Partnerships Family PD2 V V 72.1 30.7 58.6 PD4 V V 72.1 30.7 58.6 PD4 V V 92.9 92.5 58.6 Community Partnerships FAMILV1 V V 92.8 99.0 99.6 FAMILV3 V V 98.8 99.0 99.6 77.8 FAMILV3 V V 98.1 89.9 95.4 FAMILV4 V V 98.1 89.9 95.4 FAMILV5 V 83.9 na na na Community Portnerships FAMILV6 V V 98.1 89.9 95.4 Administration and Management Administration and Administration and Administration and ADMIN3 V 91.9 na na		Qualifications	STAFF8	V		20.3	na	na
			STAFF9	V		12.7	na	na
Professional Development PD1 V V 7.29 62.9 69.7 PD2 V V 75.7 72.9 74.8 PD3 V V 75.7 72.9 74.8 PD4 V V 72.1 30.7 58.6 PD4 V V 61.0 62.0 49.8 Community FAMILV2 V V 99.8 99.0 99.6 Family and FAMILV2 V V 99.8 99.0 99.6 Family and FAMILV2 V V 98.1 89.9 95.4 FAMILV4 V V 98.1 89.9 95.4 7.8 FAMILV5 V 83.9 na na na na Community FAMILV1 V V 98.2 90.7 95.7 Partnerships COMMUNITV1 V V 98.2 90.7 98.4 Management Administration </td <td></td> <td></td> <td>STAFF10</td> <td>V</td> <td></td> <td>20.2</td> <td>na</td> <td>na</td>			STAFF10	V		20.2	na	na
Development PD2 V V V 72.7 72.9 74.8 PD3 V V V 72.1 30.7 58.6 PD4 V V 61.0 27.0 49.8 Community Family Partnerships FAMILY2 V V 99.4 99.0 99.6 Partnerships FAMILY2 V V 98.1 89.9 99.4 FAMILY5 V V 98.1 89.9 95.4 FAMILY5 V 83.9 na na FAMILY5 V 98.1 88.3 94.9 COMMUNITY1 V V 98.1 88.3 94.9 COMMUNITY2 V V 98.1 88.3 94.9 COMMUNITY2 V V 98.1 88.3 94.9 COMMUNITY2 V V 99.1 66.3 83.0 Administration and Management Administration ADMIN1 V 99.9 </td <td></td> <td>Professional</td> <td>PD1</td> <td>V</td> <td>v</td> <td>72.9</td> <td>62.9</td> <td>69.7</td>		Professional	PD1	V	v	72.9	62.9	69.7
		Development	PD2	V	V	75.7	72.9	74.8
PD4 V V 61.0 27.0 49.8 Family Community Partnerships Family Partnerships FAMILV1 V V 92.4 92.9 92.5 Partnerships FAMILV2 V V 99.8 99.0 99.6 Partnerships FAMILV3 V V 99.8 89.9 95.4 FAMILV5 V 83.9 na na na FAMILV6 V V 98.1 89.9 95.7 FAMILV5 V 83.9 na na Community FAMILV6 V 98.1 88.3 94.9 COMMUNITV1 V V 98.1 88.3 94.9 COMMUNITV2 V V 98.1 88.3 94.9 COMMUNITV2 V V 99.2 97.6 98.7 Administration and Administration ADMIN1 V 90.3 na na ADMIN1 V V 90.3			PD3	V	v	72.1	30.7	58.6
Family and Community Partnerships Family Partnerships Family FAMILY2 V V 92.4 92.9 92.5 Partnerships FAMILY2 V V 99.8 99.0 99.6 Partnerships FAMILY3 V V 98.1 89.9 95.4 FAMILY5 V V 98.1 89.9 95.4 FAMILY6 V V 98.2 90.7 95.7 FAMILY6 V 98.2 90.7 95.7 Partnerships COMMUNITY1 V V 98.1 88.3 94.9 COMMUNITY2 V V 91.1 66.3 83.0 Administration and Management ADMIN1 V V 99.2 97.6 98.7 ADMIN3 V 81.8 na na na na ADMIN3 V 91.9 na na na ADMIN5 V 92.9 95.4 99.8 Environment HEALTH1<			PD4	V	v	61.0	27.0	49.8
Community Partnerships Partnerships FAMILY2 V V 99.8 99.0 99.6 Partnerships FAMILY3 V V 94.0 44.6 77.8 FAMILY4 V V 98.1 89.9 95.4 FAMILY5 V 96.5 na na FAMILY6 V 96.5 na na FAMILY5 V 98.1 88.9 95.7 Community COMMUNITY1 V V 98.1 88.3 94.9 COMMUNITY1 V V 98.1 88.3 94.9 COMMUNITY1 V V 99.2 97.6 98.7 Administration ADMIN1 V V 90.3 na na ADMIN1 V V 90.3 na na na ADMIN3 V 91.9 na na na na ADMIN4 V V 99.9 99.4 99.8	Family and	Family	FAMILY1	V	v	92.4	92.9	92.5
$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	Community	Partnerships	FAMILY2	V	v	99.8	99.0	99.6
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	Partnerships		FAMILY3	v	V	94.0	44.6	77.8
$\begin{tabular}{ c c c c c c } \hline FAMILYS V & 83.9 & na & na \\ \hline FAMILYS V & 96.5 & na & na \\ \hline FAMILYS V & 96.5 & na & na \\ \hline FAMILYS V & V & 96.7 & 75.7 \\ \hline Partnerships & COMMUNITY2 V & V & 98.1 & 88.3 & 94.9 \\ \hline COMMUNITY3 V & V & 91.1 & 66.3 & 83.0 \\ \hline Administration and Management & Administration & ADMIN1 V & V & 99.2 & 97.6 & 98.7 \\ \hline AndMaagement & and Management & ADMIN1 V & V & 90.3 & na & na \\ \hline ADMIN3 V & 81.8 & na & na \\ \hline ADMIN3 V & 90.3 & na & na \\ \hline ADMIN4 V & 90.3 & na & na \\ \hline ADMIN5 V & 91.9 & na & na \\ \hline ADMIN5 V & 91.9 & 90.9 & 99.4 & 99.8 \\ \hline Frvironment & Physical & PHYSICAL1 V & V & 99.9 & 99.4 & 99.8 \\ \hline Environment & RATIOS1 V & V & 74.2 & 35.2 & 61.4 \\ \hline Health & HEALTH1 V & V & 92.9 & 85.5 & 90.5 \\ \hline HEALTH2 V & V & 93.5 & 42.8 & 76.9 \\ \hline HEALTH3 V & V & 93.5 & 42.8 & 76.9 \\ \hline HEALTH3 V & V & 93.2 & 94.7 & 92.4 \\ \hline CURRICULUM1 V & V & 93.9 & 88.8 & 95.6 \\ \hline CURRICULUM2 V & V & 99.9 & 88.8 & 95.6 \\ \hline CURRICULUM3 V & V & 84.4 & 25.0 & 65.0 \\ \hline CURRICULUM3 V & V & 98.9 & 88.8 & 95.6 \\ \hline CURRICULUM3 V & V & 88.4 & 63.3 & 80.2 \\ \hline CURRICULUM3 V & V & 88.4 & 63.3 & 80.2 \\ \hline CURRICULUM3 V & V & 98.9 & 88.8 & 95.6 \\ \hline CURRICULUM3 V & V & 98.9 & 88.8 & 95.6 \\ \hline CURRICULUM3 V & V & 98.9 & 88.8 & 95.6 \\ \hline CURRICULUM3 V & V & 88.4 & 63.3 & 80.2 \\ \hline CURRICULUM3 V & V & 98.9 & 88.8 & 95.6 \\ \hline CURRICULUM3 V & V & 88.4 & 63.3 & 80.2 \\ \hline CURRICULUM3 V & V & 88.4 & 63.3 & 80.2 \\ \hline CURRICULUM3 V & V & 88.4 & 63.3 & 80.2 \\ \hline CURRICULUM3 V & V & 88.4 & 63.3 & 80.2 \\ \hline CURRICULUM5 V & V & 98.5 & 13.5 & 55.9 \\ \hline SCREENING2 V & V & 97.6 & 13.5 & 55.9 \\ \hline SCREENING3 V & V & 75.6 & 12.4 & 54.9 \\ \hline SCREENING5 V & V & 75.6 & 12.4 & 54.9 \\ \hline SCREENING5 V & V & 75.6 & 12.4 & 54.9 \\ \hline SCREENING5 V & V & 75.6 & 12.4 & 54.9 \\ \hline SCREENING5 V & V & 75.6 & 12.4 & 54.9 \\ \hline SCREENING5 V & V & 98.6 & na & na \\ \hline \end{array}$			FAMILY4	V	V	98.1	89.9	95.4
FAMILY6 V 96.5 na na Community COMMUNITY1 V V 98.2 90.7 95.7 Partnerships COMMUNITY2 V V 98.1 88.3 94.9 Administration and Management Administration ADMIN1 V V 99.1 66.3 88.0 Administration and Management Administration ADMIN2 V V 96.7 77.5 90.4 ADMIN2 V V 96.7 77.5 90.4 ADMIN2 V V 96.7 77.5 90.4 ADMIN2 V V 90.3 na na ADMIN2 V V 90.3 na na ADMIN5 V 91.9 na na na Environment Physical RATIOS1 V V 92.9 85.5 90.5 Environment HEALTH1 V V 92.2 91.5 92.0			FAMILY5	V		83.9	na	na
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$			FAMILY6	V		96.5	na	na
Partnersnips COMMUNITY2 V V 98.1 88.3 94.9 Administration Management Administration ADMIN1 V V 91.1 66.3 83.0 Management and Management ADMIN1 V V 99.2 97.6 98.7 Management and Management ADMIN2 V V 96.7 77.5 90.4 ADMIN3 V 81.8 na na na ADMIN5 V 90.3 na na ADMIN5 V 99.9 99.4 99.8 Environment Physical RATIOS1 V V 74.2 35.2 61.4 Health HEALTH1 V V 99.9 85.5 90.5 Environment HEALTH2 V V 94.7 92.4 Health V V 93.5 42.8 76.9 Instruction Curriculum and Curriculum2 V V 98.9<		Community	COMMUNITY1	v	V	98.2	90.7	95.7
Administration and Management Administration and Management AdMiN1 V V 91.1 66.3 83.0 Management and Management ADMIN1 V V 99.2 97.6 98.7 Management and Management ADMIN2 V V 99.2 97.6 98.7 ADMIN3 V V V 96.7 77.5 90.4 ADMIN4 V 90.3 na na ADMIN5 V 91.9 na na ADMIN5 V 99.9 99.4 99.8 Environment Physical PHYSICAL1 V V 92.9 85.5 90.5 Environment HEALTH1 V V 92.9 85.5 90.5 HEALTH2 V V 91.2 94.7 92.4 HEALTH4 V V 92.2 91.5 92.0 HEALTH4 V V 98.7 77.8 99.1		Partnerships		v	v	98.1	88.3	94.9
Administration Admini	Adapteter	A	COMMUNITY3	v	v	91.1	66.3	83.0
	Administration and	Administration	ADMIN1	V	V	99.2	97.6	98.7
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	wanagement	and Management	ADMIN2	V	v	96.7	//.5	90.4
$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$				V		81.8	na	na
Environment Physical Environment Physical Environment PHYSICAL1 V 99.9 99.4 99.8 Ratios RATIOS1 V V 99.9 99.4 99.8 Ratios RATIOS1 V V 99.9 99.4 99.8 Health HEALTH1 V V 92.9 85.5 90.5 Environment HEALTH1 V V 91.2 94.7 92.4 HEALTH3 V V 93.5 42.8 76.9 HEALTH3 V V 93.5 42.8 76.9 HEALTH3 V V 93.5 42.8 76.9 HEALTH5 V V 45.7 70.2 53.7 Curriculum and Instruction CURRICULUM1 V V 98.9 88.8 95.6 CURRICULUM3 V V 88.4 63.3 80.2 CURRICULUM3 V V 88.6 39.7 72.6 SC				V		90.3	na	na
$\begin{tabular}{ c c c c c c c } \hline PH SCALL V V 95.5 9.5.4 95.4 95.8 \\ \hline Environment \\ \hline Ratios RATIOS1 V V 7 74.2 35.2 61.4 \\ \hline Health Health HEALTH1 V V 92.9 85.5 90.5 \\ \hline Environment \\ \hline HEALTH2 V V 91.2 94.7 92.4 \\ \hline HEALTH3 V V 93.5 42.8 76.9 \\ \hline HEALTH3 V V 92.2 91.5 92.0 \\ \hline HEALTH5 V V 45.7 70.2 53.7 \\ \hline Curriculum and Instruction \\ \hline CURRICULUM1 V V 98.9 88.8 95.6 \\ \hline CURRICULUM2 V V 99.7 97.8 99.1 \\ \hline CURRICULUM3 V V 99.7 97.8 99.1 \\ \hline CURRICULUM3 V V 88.4 63.3 80.2 \\ \hline CURRICULUM4 V V 90.2 65.1 82.0 \\ \hline CURRICULUM5 V V 90.2 65.1 82.0 \\ \hline SCREENING1 V V 88.6 39.7 72.6 \\ \hline SCREENING2 V V 88.2 38.3 69.8 \\ \hline SCREENING3 V V 76.6 13.5 55.9 \\ \hline SCREENING5 V V 75.6 12.4 54.9 \\ \hline SCREENING5 V V 75.6 12.4 54.9 \\ \hline SCREENING5 V V 75.6 12.4 54.9 \\ \hline SCREENING5 V V 772.8 12.8 53.2 \\ \hline Consistent CONSISTENT1 V 98.6 na na \\ \hline \end{tabular}$	Environmont	Dhysical		v	2/	91.9	00.4	
$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	Environment	Environment	PHISICALI	v	v	99.9	99.4	99.0
$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$		Batios	RATIOS1	N	٧	74.2	35.2	61.4
Inclum Inclum<		Health	HFAITH1	v	v v	92.9	85.5	90.5
HEALTH2 V V 93.5 42.8 76.9 HEALTH3 V V 93.5 42.8 76.9 HEALTH4 V V 92.2 91.5 92.0 HEALTH4 V V 92.2 91.5 92.0 HEALTH5 V V 45.7 70.2 53.7 Curriculum and Instruction Curriculum CURRICULUM1 V V 98.9 88.8 95.6 CURRICULUM2 V V 99.7 97.8 99.1 CURRICULUM3 V V 88.4 63.3 80.2 CURRICULUM3 V V 88.4 63.3 80.2 CURRICULUM4 V V 88.6 39.7 72.6 SCREENING2 V V 88.6 39.7 72.6 SCREENING3 V V 76.6 13.5 55.9 SCREENING3 V V 75.6 12.4 54.9 <t< td=""><td></td><td>Environment</td><td>HEALTH2</td><td>v</td><td><u>ب</u></td><td>91.2</td><td>94.7</td><td>92.4</td></t<>		Environment	HEALTH2	v	<u>ب</u>	91.2	94.7	92.4
HEALTHA V V 92.2 91.5 92.0 HEALTH4 V V 92.2 91.5 92.0 HEALTH5 V V 45.7 70.2 53.7 Curriculum and Instruction Curriculum CURRICULUM1 V V 98.9 88.8 95.6 CURRICULUM2 V V 99.7 97.8 99.1 CURRICULUM3 V V 88.4 63.3 80.2 CURRICULUM3 V V 88.4 63.3 80.2 CURRICULUM4 V V 88.4 63.3 80.2 CURRICULUM5 V V 90.2 65.1 82.0 Assessment SCREENING1 V V 88.6 39.7 72.6 SCREENING2 V V 85.2 38.3 69.8 SCREENING3 V V 75.6 12.4 54.9 SCREENING5 V V 72.8 12.8 53.2			HFAITH3	v	v	93.5	42.8	76.9
Curriculum and Instruction Curriculum Curriculum CURRICULUM1 V V 45.7 70.2 53.7 Curriculum and Instruction Curriculum CURRICULUM1 V V 98.9 88.8 95.6 CURRICULUM2 V V 99.7 97.8 99.1 CURRICULUM3 V V 88.4 63.3 80.2 CURRICULUM4 V V 88.4 63.3 80.2 CURRICULUM5 V V 90.2 65.1 82.0 Assessment SCREENING1 V V 88.6 39.7 72.6 SCREENING2 V V 88.6 39.7 72.6 SCREENING3 V V 76.6 13.5 55.9 SCREENING3 V V 76.6 13.5 53.2 SCREENING5 V V 72.8 12.4 54.9 SCREENING5 V V 72.8 12.8 53.2 Consi			HEALTH4	V	v	92.2	91.5	92.0
Curriculum and Instruction Curriculum CURRICULUM1 V V 98.9 88.8 95.6 CURRICULUM2 V V 99.7 97.8 99.1 CURRICULUM3 V V 88.4 63.3 80.2 CURRICULUM4 V V 88.6 39.7 72.6 Screening1 V V 88.6 39.7 72.6 Screening2 V V 85.2 38.3 69.8 Screening3 V V 75.6 12.4 54.9 Screening5 V V 72.8 12.8 53.2 Consistent CONSISTENT1 V 98.6 na na			HEALTHS	V	۰ ۷	45.7	70.2	53.7
Instruction CURRICULUM2 V V 99.7 97.8 99.1 CURRICULUM3 V V 84.4 25.0 65.0 CURRICULUM4 V V 88.4 63.3 80.2 CURRICULUM4 V V 88.4 63.3 80.2 CURRICULUM5 V V 90.2 65.1 82.0 Assessment SCREENING1 V V 88.6 39.7 72.6 SCREENING2 V V 85.2 38.3 69.8 SCREENING3 V V 76.6 13.5 55.9 SCREENING4 V V 72.6 12.4 54.9 SCREENING5 V V 72.6 12.4 54.9 SCREENING5 V V 72.8 12.8 53.2 Consistent Caregiver CONSISTENT1 V 98.6 na na	Curriculum and	Curriculum	CURRICULUM1	V	v	98.9	88.8	95.6
CURRICULUM3 V V 84.4 25.0 65.0 CURRICULUM4 V V 88.4 63.3 80.2 CURRICULUM5 V V 90.2 65.1 82.0 Assessment SCREENING1 V V 88.6 39.7 72.6 SCREENING2 V V 85.2 38.3 69.8 SCREENING3 V V 76.6 13.5 55.9 SCREENING4 V V 75.6 12.4 54.9 SCREENING5 V V 72.8 12.8 53.2 Consistent Caregiver CONSISTENT1 V 98.6 na na	Instruction	Carriedan	CURRICULUM2	V	۰ ۷	99.7	97.8	99.1
CURRICULUM4 V V 88.4 63.3 80.2 CURRICULUM5 V V 90.2 65.1 82.0 Assessment SCREENING1 V V 88.6 39.7 72.6 SCREENING2 V V 88.6 39.7 72.6 SCREENING2 V V 85.2 38.3 69.8 SCREENING3 V V 76.6 13.5 55.9 SCREENING4 V V 75.6 12.4 54.9 SCREENING5 V V 72.8 12.8 53.2 Consistent Caregiver CONSISTENT1 V 98.6 na na			CURRICULUM3	v	٧	84.4	25.0	65.0
CURRICULUM5 V V 90.2 65.1 82.0 Assessment SCREENING1 V V 88.6 39.7 72.6 SCREENING2 V V 88.6 39.7 72.6 SCREENING2 V V 85.2 38.3 69.8 SCREENING3 V V 76.6 13.5 55.9 SCREENING4 V V 75.6 12.4 54.9 SCREENING5 V V 72.8 12.8 53.2 Consistent Caregiver CONSISTENT1 V 98.6 na na			CURRICULUM4	v	٧	88.4	63.3	80.2
Assessment SCREENING1 V V 88.6 39.7 72.6 SCREENING2 V V 85.2 38.3 69.8 SCREENING3 V V 76.6 13.5 55.9 SCREENING4 V V 75.6 12.4 54.9 SCREENING5 V V 72.6 12.8 53.2 Consistent Caregiver CONSISTENT1 V 98.6 na na			CURRICULUM5	v	٧	90.2	65.1	82.0
SCREENING2 V V 85.2 38.3 69.8 SCREENING3 V V 76.6 13.5 55.9 SCREENING4 V V 75.6 12.4 54.9 SCREENING5 V V 72.8 12.8 53.2 Consistent Caregiver CONSISTENT1 V 98.6 na na		Assessment	SCREENING1	v	V	88.6	39.7	72.6
SCREENING3 V V 76.6 13.5 55.9 SCREENING4 V V 75.6 12.4 54.9 SCREENING5 V V 72.8 12.8 53.2 Consistent Caregiver CONSISTENT1 V 98.6 na na			SCREENING2	٧	٧	85.2	38.3	69.8
SCREENING4 V V 75.6 12.4 54.9 SCREENING5 V V 72.8 12.8 53.2 Consistent CONSISTENT1 V 98.6 na na Caregiver Consistent <			SCREENING3	٧	٧	76.6	13.5	55.9
SCREENING5VV72.812.853.2ConsistentCONSISTENT1V98.6nanaCaregiverCaregiverCaregiverCaregiverCaregiverCaregiver			SCREENING4	٧	٧	75.6	12.4	54.9
Consistent CONSISTENT1 V 98.6 na na Caregiver			SCREENING5	٧	٧	72.8	12.8	53.2
Caregiver		Consistent	CONSISTENT1	٧		98.6	na	na
		Caregiver						

Note. For the shared items, response rate was calculated based on combined data.

Exploratory Factor Analysis (EFA) results did not support a five-factor structure.

Acknowledging the problem with the data — that is, they contained several "easy items" and "difficult items" that could affect the integrity of the results — the research team decided to conduct exploratory factor analysis on the center- and home-combined data, which showed a higher level of variation across all items. The research team also ran EFAs on the center-based sample and the home-based sample separately, exploring the theme for a potential factor structure.

Center- and home-combined data did not support a five-factor structure.

We conducted EFAs on the combined center-based and home-based samples. Because all the items are on a binary scale, we used a tetrachoric correlation matrix and full information maximum likelihood estimation to explore the factor structure using Mplus 7.4 (Muthen & Muthen, 2010). First, we calculated eigenvalues to examine the number of potential underlying factors. Because some items were not administered in the home-based sample, performing maximum likelihood estimates or robust estimates on all items, including these center-only items, would contradict the original design of the home-based survey. Therefore, only items that were universal to both types of classrooms were included in the analysis. Figure A1 presents the eigenvalues of the exploratory factor analysis in a scree plot. To determine the number of factors, the research team applied an empirical cutoff value of 1. Factors with eigenvalues less than 1 were discarded. The graph indicates that no more than eight factors should be retained.





Note. Model with negative eigenvalues indicates a problematic data structure, which may be due to highly skewed item response.

The research team tested the extent to which models with one to five factors fit the data. We did not go beyond five factors because the survey was originally designed to measure five domains of program quality. Creating more factors than the original design would go against the design as well as the purpose of item reduction. We performed geomin and varimax rotations to obtain the optimal factor structure. Geomin rotation is a type of oblique rotation wherein all underlying factors are allowed to correlate. On the contrary, varimax rotation is a type of orthogonal rotation wherein all factors are forced to be uncorrelated. Results showed that the covariate matrix for varimax rotations was not positively definite, indicating that data were not structured in a way that can be analyzed in a varimax rotation. Therefore,

the research team reported only findings with geomin rotation. In addition, model goodness of fit criteria (comparative fit index [CFI] and Tucker-Lewis index [TLI] larger than 0.90, root mean square error of approximation [RMSEA] and standardized root mean residual [SRMR] smaller than 0.08–0.10, optimally 0.05), factor loadings with a threshold value of 0.3, double loaders, and the meaningfulness of each factor were considered when selecting the final model. As shown in Table A2, a five-factor model was accepted. In addition, two- to four-factor models reached a marginal level of acceptance.

Number of Factors	CFI	TLI	RMSEA	SRMR	Model Acceptance
1 factor	0.90	0.90	0.07	0.14	Rejected
2 factors	0.95	0.95	0.05	0.10	Accept with reservation
3 factors	0.97	0.96	0.04	0.08	Accept with reservation
4 factors	0.98	0.97	0.04	0.07	Accept with reservation
5 factors	0.99	0.98	0.03	0.06	Accepted

Table A2. Determining the Goodness of Fit for Different Factor Structures (Combined)

Note. CFI is the Comparative fit index; TLI is the Tucker Lewis index; RMSEA is the Root mean square error of approximation; SRMR is the Standardized root mean square residual.

We next examined the five-factor, four-factor, three-factor, and two-factor models using the goemin rotation to examine the meaningfulness of each factor that emerged from these structures (e.g., we examined whether items grouped together in meaningful way or if the patterns random). Before looking at the factors, we removed items that were double loaders (e.g., associated with two or more factors) or that did not load on any factor. With these analyses, we determined that a five-factor structure fit the data best;⁶ however, this five-factor structure did not align with the five domains of quality as intended. First, thirteen items did not load on any factor or double loaded. These items did not indicate a unique factor (or topic area) and we recommend excluding them in subsequent analyses. While we found a five-factor structure to be the best fitting, the remaining 25 items loaded primarily on three factors, which we identified as Staff Qualifications, General Program Operation, and Screening.

⁶ The four-factor model excluded 16 items, and most of the remaining 22 items loaded on one factor out of the four. The three-factor model excluded 16 items, and the remaining 22 items loaded mostly on one factor. Finally, the two-factor model excluded 11 items, but the remaining 27 items did not form meaningful factors. Taken together, the combined data may not support a meaningful factor structure.

Hypothesized Domain	Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
	DIRECTOR2	0.809				
	STAFF2	0.650				
Professional Development	STAFF3	0.671				
	STAFF4	0.373				
	PD2				0.612	
	FAMILY2			0.760		
	FAMILY3		0.900			
	FAMILY4			0.498		
Family and Administration	COMMUNITY1		0.778			
	COMMUNITY2		0.867			
	ADMIN1		0.514			
	ADMIN2	0.409				
	RATIOS1	0.750				
	HEALTH1			0.440		
Environment	HEALTH2			0.410		
	HEALTH3	0.594				
	HEALTH5				0.485	
	CURRICULUM1			0.680		
	CURRICULUM2			0.754		
	CURRICULUM3	0.908				
	SCREENING1	0.845				
Curriculum	SCREENING2	0.775				
	SCREENING3	0.953				
	SCREENING4	0.984				
	SCREENING5	1.010				
Items recommended to						
exclude						
	DIRECTOR1	0.002	-0.639	-0.128	0.608	0.288
	DIRECTOR3	0.459	0.167	-0.011	-0.414	0.418
	DIRECTOR4	-0.359	0.804	-0.037	-0.005	-0.626
Staff Qualifications and	STAFF1	0.056	-0.633	-0.063	0.561	0.233
Professional Development	PD1	0.010	0.514	0.052	0.632	-0.075
	PD3	-0.219	0.456	0.339	-0.095	0.148
	PD4	-0.049	0.238	0.443	-0.087	0.372
Family and Community	FAMILY1	-0.038	-0.011	0.505	0.536	-0.316
Collaboration	COMMUNITY3	0.006	0.323	0.407	-0.006	0.127
	HEALTH4	0.117	0.034	0.430	0.518	-0.516
Environment	PHYSICAL1	-0.302	0.046	0.473	-0.032	-0.032
	CURRICULUM4	0.032	0.014	0.777	-0.011	0.361
Curriculum	CURRICULUM5	-0.023	0.027	0.835	0.012	0.370

Table A3. Factor Loadings for the Five-Factor Model (Combined)

Note. Analysis was based on center- and home-combined data. Double-loaded items and items that did not load on any factors were moved to the bottom of the table. The tetrachoric correlation matrix was semidefinite, which indicates an issue with the data. The problem could be due to the highly skewed distribution of some items. Interpretation of the results therefore should be made with caution. Factor loading below .30 are not meaningful and are excluded from the table for clarity.

N = 4,397.

Center-based program data suggested two emerging factors.

While we focused on the findings for the combined sample for our main findings, we also ran EFAs on the center-based and home-based samples to explore how the factors within these samples were similar or different from the combined sample. Differences in these structures may indicate that the SAS is functioning in different ways for one sample or another, which could be problematic if we consider the GSQ ratings comparable across program types. Figure A2 presents the eigenvalues of the EFA in a scree plot for the center-based sample. The graph indicates that no more than 14 factors should be retained. As with the combined sample, we focused these analyses on the one- to five-factor analyses to be consistent with the five-quality domain structure of the SAS protocols.



Figure A2. Eigenvalues of the Exploratory Factor Analysis, SAS (Center-Based)

Note. Model with negative eigenvalues indicated a problematic data structure, which may be due to highly skewed item response.

The research team tested the extent to which models with one to five factors fit the data. As shown in Table C4, only a five-factor model reached a marginal level of acceptance (with SRMR = 0.10).

Number of Factors	CFI	TLI	RMSEA	SRMR	Model Acceptance
1 factor	0.81	0.80	0.06	0.17	Rejected
2 factors	0.89	0.88	0.05	0.13	Rejected
3 factors	0.94	0.93	0.04	0.12	Rejected
4 factors	0.96	0.95	0.03	0.11	Rejected
5 factors	0.96	0.96	0.03	0.10	Accept with reservation

Table A4. Determining the Goodness of Fit for Different Factor Structures (Center-Based)

Note. CFI is the Comparative fit index; TLI is the Tucker Lewis index; RMSEA is the Root mean square error of approximation; SRMR is the Standardized root mean square residual.

After excluding double loaders and items that did not load on any factors (23 items), 27 items remained with two prominent factors: Staff Qualifications (factor 3) and General Program Operations (factor 2). No items loaded on factor 1, factor 4 had only one item, and factor 5 had only two items. Because these items did not construct a meaningful five-factor model, this model should be rejected. Table A5 presents the factors and item loadings.

Hypothesized Domain	Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
	DIRECTOR1			0.633		
	STAFF1			0.823		
	STAFF2			0.746		
	STAFF3			0.694		
Staff and Professional Development	STAFF4			0.435		
	STAFF5			0.645		
	STAFF7			0.608		
	STAFF8		0.456			
	FAMILY1					0.727
	FAMILY2				0.990	
	FAMILY3	0.669				
	FAMILY4	0.738				
Family and Administration	FAMILY5	0.697				
	FAMILY6	0.653				
	COMMUNITY1	0.793				
	COMMUNITY3	0.682				
	ADMIN2	0.509				
	ADMIN3	0.519				
Administration	ADMIN4	0.603				
	ADMIN5	0.535				
	HEALTH1	0.654				
Environment	HEALTH3	0.478				
	HEALTH4					0.991
	CURRICULUM1	0.619				
Control of	CURRICULUM4	0.700				
Curriculum	CURRICULUM5	0.752				
	SCREENING2	0.721				
Items recommended to exclude						
	DIRECTOR2	-0.491	-0.14	0.408	0.078	0.063
	DIRECTOR3	-0.239	0.320	0.197	-0.156	-0.611
	DIRECTOR4	0.966	-0.029	-0.029	0.018	0.728
	STAFF6	0.294	0.305	-0.497	-0.406	-0.093
Staff Qualifications and Professional	STAFF9	0.200	0.218	0.217	-0.009	0.100
Development	STAFF10	-0.089	0.055	-0.465	-0.629	0.338
	PD1	0.029	0.340	0.182	-0.191	0.634
	PD2	-0.013	0.357	0.220	-0.063	0.594
	PD3	0.464	0.545	0.091	-0.02	-0.056
	PD4	0.064	0.670	0.253	-0.001	-0.328
Family and Community Collaboration	COMMUNITY2	-0.127	0.805	-0.017	0.335	0.043
Administration	ADMIN1	-0.033	0.503	-0.174	0.304	0.002
	PHYSICAL1	0.079	0.051	-0.937	0.659	0.161
Environment	RATIOS1	0.198	0.391	-0.172	-0.464	-0.082
Livionnent	HEALTH2	-0.017	0.221	-0.070	-0.017	0.269
	HEALTH5	-0.313	0.462	-0.125	-0.035	0.317
	CURRICULUM2	-0.021	-0.086	-0.386	0.380	0.439
	CURRICULUM3	0.124	0.676	-0.076	-0.417	0.022
	SCREENING1	-0.005	0.595	0.052	-0.163	0.457
Curriculum	SCREENING3	0.074	0.837	-0.016	-0.412	0.006
	SCREENING4	0.097	0.833	0.007	-0.469	0.060
	SCREENING5	0.037	0.732	0.022	-0.480	0.373
	CONSISTENT1	0.286	0.255	-0.137	0.183	0.112

Table A5. Factor Loadings for the Five-Factor Model (Center-Based)

Note. Analysis was based on center-only data. Double-loaded items and items that did not load on any factors were moved to the bottom of the table. The tetrachoric correlation matrix was semidefinite, which indicates an issue with the data. The problem could be due to the highly skewed distribution of some items. Interpretation of the results should be made with caution. Factor loadings below .30 are not meaningful and are excluded from the table for clarity. N = 2,956.

Home-based program data suggested a two-factor structure.

Finally, we performed an EFA on the home-based sample. Figure A3 presents the eigenvalues of the EFA in the format of a scree plot for home-based classrooms. Figure A3 presents the eigenvalues of the EFA in the format of a scree plot for the home-based sample. The graph indicates that no more than nine factors should be retained. As with the combined and center-based samples we focused these analyses on the one to five factor analyses to be consistent with the five quality domain structure of the SAS protocols.





The research team tested the extent to which models with one to five factors fitted the data.. As shown in Table A6, a three-factor model and a four-factor model reached a marginal level of acceptance (SRMR under 0.10). A five-factor model was rejected because the model was unidentified, because the covariance matrix included many extreme values. This again was related to the highly skewed items on the SAS where over 95% of programs responded "yes" to an item, or had achieved that aspect of program quality. With the presence of these extreme values in the survey and the covariance matrix, the mathematical analysis could not compute standard errors which are essential to estimate all of the model's parameters, creating a statistical error, or an "unidentified" model. As such, these analyses should be interpreted with extreme caution, and most likely rejected.

Number of Factors	CFI	TLI	RMSEA	SRMR	Model Acceptance
1 factor	0.84	0.83	0.08	0.18	Rejected
2 factors	0.92	0.91	0.05	0.13	Rejected
3 factors	0.95	0.95	0.04	0.10	Accept with reservation
4 factors	0.97	0.96	0.04	0.09	Accept with reservation
5 factors	*	*	*	*	Rejected

Table A6. Determining the Goodness of Fit for Different Factor Structures (Home-Based)

Note. * = Could not compute standard errors due to severe problem with covariance matrix. Model unidentified. CFI: Comparative is the fit index; TLI is the Tucker Lewis index; RMSEA is the Root mean square error of approximation; SRMR is the Standardized root mean square residual.

Note. Model with negative eigenvalues indicates a problematic data structure, which may be due to highly skewed item response.

After excluding double loaders and items that did not load on any factors (10 items), 27 items loaded on three out of four factors (see Table A7). The three-factor model excluded 18 items, with the remaining 19 items loaded on two factors. Finally, a two-factor model was examined. Eight items were removed, and 29 items were retained to represent Staff Qualifications (factor 1) and General Program Operations (factor 2). However, because the model-fit statistics for the two-factor model were marginally acceptable, interpretations of the results should be made carefully.

	ltem	Factor 1	Factor 2
	DIRECTOR1	-0.762	
	DIRECTOR2	0.970	
	DIRECTOR3	0.624	
	DIRECTOR4	0.592	
	STAFF1	-0.755	
Staff Qualifications and	STAFF2	0.976	
Professional Development	STAFF3	0.618	
	STAFF4	0.589	
	PD1		0.528
	PD2		0.559
	PD3		0.744
	PD4		0.799
	FAMILY1		0.635
	FAMILY2		0.63
The set is a set of the state is a state of	FAMILY4		0.535
Family and Administration	COMMUNITY1		0.774
	COMMUNITY2		0.769
	COMMUNITY3		0.448
	ADMIN1		0.550
Administration	ADMIN2		0.332
	HEALTH1		0.579
	HEALTH2		0.532
Environment	HEALTH3		0.374
	HEALTH4		0.577
	HEALTH5		0.472
	CURRICULUM1		0.593
	CURRICULUM2		0.765
Curriculum	CURRICULUM4		0.853
	CURRICULUM5		0.89
Items recommended to be exclude:			
Family and Community Collaboration	FAMILY3	0.479	0.509
Environment	RATIOS1	0.239	0.052
	CURRICULUM2	0.580	0.484
	SCREENING1	0.511	0.541
Curriculum	SCREENING2	0.439	0.605
Curriculum	SCREENING3	0.601	0.598
	SCREENING4	0.589	0.625
	SCREENING5	0.600	0.608

Table A7. Factor Loadings for the Two-Factor Model (Home-Based)

Note. Analysis was based on home-based data. Double-loaded items and items that did not load on any factors were moved to the bottom of the table. The tetrachoric correlation matrix was semidefinite, which indicates an issue with the data. The problem could be due to the highly skewed distribution of some items. Interpretation of the results should be made with caution. PHYSICAL1 was forced to be excluded from the analysis by the statistical software, because 100% of the home-based classrooms selected "Yes" on this item. Factor loading below .30 are not meaningful and are excluded from the table for clarity. N = 1,441.

Rasch Analysis

For the Rasch analysis, the research team used an aggregate dataset of all individual SAS responses regardless of protocol type. The file includes the data as reported by individual respondents (rather than program-level averages or a validated score) to test the properties of how well the tool is working as a self-report measure. Whenever possible, we combined indicators measuring the same concept across the three protocols. In the case of an item asked on only one protocol, the item was left missing for all other respondents (which the Rasch analysis can adjust for). Two types of variables were used in thisfile:

- *Categorical Variables:* For the director and staff education indicators, we recoded the set of indicators to a categorical variable ranging from 0 to 3 or 4, depending on the number of score categories in the SAS protocols.
- *Binary Indicator Variables:* All other variables were recoded as 0 or 1 to indicate whether or not an item was present.

We conducted a series of partial-credit Rasch models (Wright & Masters, 1982) to the test the properties of the following:

- The five intended factors of the SAS on the combined sample;
- A single factor, including all items on the combined sample;
- A single factor, including all items in the center-based sample; and
- A single factor, including all items on the home-based sample.

The partial-credit Rasch model assigns an overall, quantitative measure to each individual on a given factor, based on their responses to one or more items within the construct and the difficulty of each item. When items function well together, clear patterns emerge from which we can derive the difficulty of each item within a factor, as well as the strength of endorsement of each individual on the factor. For example, in a well-functioning factor, we observe survey participants consistently agreeing more strongly with one item and less strongly with another item. With multiple items within a factors, we can then estimate each individual's "score" on the construct based on his or her responses relative to other survey respondents.

Before calculating the overall "score," we examined the internal consistency and reliability of each construct to determine if the constructs were functioning as intended. First, we examined two measures of reliability, person separation reliability and Cronbach's alpha, measures of how consistently survey respondents answered each question relative to their other responses as well as to one another. For both measures, reliability ranges from 0 (no reliability) to 1 (perfect reliability) and typically scores above 0.70 are considered good, and scores above 0.80 are considered to be even better.

Next, we examined the outfit and point-correlation range for each construct. *Outfit* measures the extent to which the observed ratings are aligned with expected ratings based on the ratings on all other items (Bond & Fox, 2007). An outfit of 1.00 indicated a perfect fit of an item with the construct, and an outfit between 0.5 and 1.5 was considered good fitting. Finally, we examined the correlation of each item to the overall construct.

The findings from the five intended factors of the SAS on the combined sample are presented in Chapter 3. In the sections that follow, we present more details from the single-factor analyses on the combined, center-based, and home-based samples.

The SAS reliably measures program quality; however, further revisions could be made to improve the measure as a whole.

The SAS as a single-factor measure (before any revisions) functioned reliably for the center, homes, and for the overall combined sample. The reliability measures were 0.70 or above. However, there is evidence of items that do not fit the data, as well as items that did not correlate strongly with the overall SAS scores. These items are candidates for removal in a streamlined version of the SAS.

Construct	Number of Self- Assessment Survey (SAS) Records	Number of Items	Person Separation Reliability	Internal Consistency (Cronbach's Alpha)	Range of Item Fit	Point- Correlation Range
Center-Only Sample	2,956	39	0.76	0.84	0.45–2.19	0.02–0.68
Family-Only Sample	1,441	36	0.82	0.81	0.42–3.56	0.10–0.57
Combined Sample	4,397	43ª	0.86	0.92	0.44–2.56	0.07–0.78

Table A8. Construct Reliability and Properties

^aThe combined file includes all possible SAS items across the three protocols; therefore, the number of possible items exceeds the number of items on any single SAS protocol.

Table A9 also shares all items on the SAS and their associated "difficulty."

Item Name	Respondent Sample	%age "Yes"
PHYSICAL1	All programs	< 100%
FAMILY2	All programs	< 100%
CURRICULUM2	All programs	99%
ADMIN_home3	Homes without assistants only	97%
ADMIN_home1	Homes only	97%
CONSISTENT1	Centers only	99%
CURRICULUM1	All programs	96%
FAMILY4	All programs	95%
COMMUNITY1	All programs	96%
COMMUNITY2	All programs	95%
ADMIN1	Centers and Homes with Assistants only	99%
HEALTH2	All programs	92%
FAMILY1	All programs	93%
FAMILY6	Centers only	97%
ADMIN2	Centers only	90%
HEALTH4	All programs	92%
HEALTH1	All programs	91%
ADMIN_home4	Homes without assistants only	76%
ADMIN5	Centers only	92%
COMMUNITY3	All programs	83%
CURRICULUM5	All programs	82%
ADMIN_home2	Homes only	67%
CURRICULUM4	All programs	80%
ADMIN4	Centers only	90%
FAMILY3	All programs	78%
HEALTH3	All programs	77%
DIRECTOR_education	All programs	NA
PD2	All programs	75%
SCREENING1	All programs	73%
FAMILY5	Centers only	84%
STAFF_education	All programs	NA
ADMIN3	Centers only	82%
SCREENING2	All programs	70%
PD1	All programs	70%
CURRICULUM3	All programs	65%
RATIOS1	All programs	61%
PD3	All programs	59%
SCREENING3	All programs	56%
SCREENING4	All programs	55%
HEALTH5	All programs	54%
SCREENING5	All programs	53%
PD4	All programs	50%
STARE Assistants Education	Contors and Homos with Assistants only	NA

Table A9. SAS Items Listed From Easiest to Hardest (Combined Sample)

There were also key items that typically did not fit well with the overall measure. For four items, over 98% of all survey participants responded "Yes" to the item; see Table A10 for a list of these items. On these items, we expect even the lowest quality program to respond "Yes" to the item; therefore, the item does not contribute to the measure's ability to differentiate quality.

Domain	Subdomain	Variable	Center- Based Measure	Home- Based Measure	Combined	"Yes" Response
Family and Community Partnerships	Family Partnerships and Family Strengthening	FAMILY2	V		V	99.1%
Administration and Management	Administration and Management	ADMIN1	V			98.1%
Environment	Physical	PHYSICAL1	V	V	v	99.9%
Curriculum and Instruction	Curriculum	CURRICULUM2	٧			98.4%

Table A10. Easy Items That Do Not Differentiate Quality in Any Programs by Sample Type

Participation rates in Great Start to Quality (GSQ)

Overall, 41% of all licensed early childhood programs participated in the GSQ and Table B1 provides detailed descriptive statistics for participation rates by program background characteristics. Participation was relatively even across the state of Michigan and Table B2 provides detailed participation rates by region, with participation rates ranging from 37 to 59 %. Table B3 then compares participation rates by program type and background characteristics, for further detail.

	All I	Programs	Participat	ing Programs	
		%age		%age	%age of All
Characteristic	Ν	(by row)	N	(by row)	Programs
Program Type					
Licensed child care center	4,475	48.5%	2,118	55.4%	47.3%
Registered family child care home	1,731	18.7%	692	18.1%	40.0%
Licensed group child care home	3,026	32.7%	1,011	26.5%	33.4%
Alternative Pathways					
GSRP	1,164	13.0%	1,090	28.5%	93.6%
Head Start	531	5.7%	484	12.7%	91.1%
Early Head Start	200	2.1%	164	4.3%	82.0%
Education approach					
Montessori	177	1.9%	73	1.9%	41.2%
Reggio-inspired	115	1.2%	68	1.8%	59.1%
Religious	639	6.9%	295	7.7%	46.2%
Accreditation					
NAEYC	180	1.9%	137	3.6%	76.1%
NAFCC	78	1.0%	49	1.3%	62.8%
Ages Served					
Preschool	3,528	38.2%	2,206	57.7%	62.5%
Infants and toddlers	3,386	36.6%	2,537	66.4%	74.9%
		Mean		Mean	
	N	(SD)	Ν	(SD)	
		39.19		44.14	
License Capacity		(49.47)		(52.54)	
Total	9,243		3,821		41.3%

Table B1. Comparison of Participation in GSQ by Key Program Characteristic

Note: MI QRIS Data; N = 9,243

The MI QRIS dataset has 9,243 unique license identification numbers. However, 1,237 programs are missing data on alternative pathways (GSRP, Head Start, Early Head Start), education approach (Montessori, Reggio-inspired, Religious), accreditation status (NAEYC, NAFCC), and ages served (preschool, toddlers).

SD is standard deviation, GSRP is Great Start to Readiness Program, NAEYC is National Association for the Education of Young Children, and NAFCC is National Association for Family Child Care

Table B2. Comparison of All Michigan Early Childhood Programs and Programs Participating in GSQ, by Region

	Total N in a Region	N of Programs Participating in the QRIS in a Region	%age of the ECE Programs Within a Given Region That Participate in the QRIS
Central Region	880	350	39.8%
Eastern Region	610	294	48.2%
Kent Region	682	360	52.8%
Northeast Region	229	144	62.9%
Northwest Region	349	165	47.3%
Oakland-Macomb Region	1,207	482	39.9%
Southeast Region	1,103	525	47.6%
Southwest Region	965	441	45.7%
Upper Peninsula Region	318	186	58.5%
Wayne Region	928	486	52.4%
Western Region	735	388	52.8%
Total	8,006 ¹	3,821	

Note. Michigan QRIS data .

1. 1,237 of the total 9,243 programs are missing data on region.

QRIS is Quality Rating and Improvement System and ECE is early childhood education.

Table B3. Comparison of Programs Participating in GSQ, by Program Type and Other KeyCharacteristics

	Child Care	Crown Homo	Family Llama	Ν
	Center	Group Home	Family Home	IN
Alternative Pathways				
GSRP	98.8%	0.1%	1.1%	1090
Head Start	98.4%	0.4%	1.2%	484
Early Head Start	79.3%	6.7%	14.0%	164
Educational Approach				
Montessori	68.5%	11.0%	20.5%	73
Reggio	64.7%	14.7%	20.0%	68
Religious	68.1%	12.5%	19.3%	295
Accreditation				
NAEYC	99.3%	0.7%	0.0%	137
NAFCC	0.0%	55.1%	44.9%	49
Infants & Toddlers				
Yes	34.0%	27.0%	39.0%	2,537
No	98.0%	0.3%	1.6%	1,278
Preschool				
Yes	78.0%	11.6%	10.4%	2,206
No	24.5%	27.1%	48.4%	1,615
Schedule				
Traditional	57.0%	17.8.0%	25.2%	2,640
Non-traditional	51.9%	18.7%	29.4%	1,181

Note. Michigan QRIS data; N = 3,821.

GSRP is Great Start to Readiness Program, NAEYC is National Association for the Education of Young Children, and NAFCC is National Association for Family Child Care

Detailed Findings on the Distribution of GSQ and SAS ratings

As described in the main body, the most common GSQ rating was a 3-star rating. The same was true for the published GSQ rating, see Table B4. The ratings varied by program characteristics as well, presenting in Table B5.

Rating	N	Mean (SD)	1-star n (%)	2-star n (%)	3-star <i>n</i> (%)	4-star <i>n</i> (%)	5-star <i>n</i> (%)
SAS Rating	3,821	3.56 (1.15)	127 (3.32)	450 (11.78)	1,623 (42.48)	381 (9.97)	1,240 (32.45)
Published GSQ Rating	3,796	3.27 (0.87)	127 (3.35)	448 (11.80)	1,693 (44.60)	1,314 (34.62)	214 (5.64)

Table B4. Sample Comparison of SAS and Published GSQ Ratings

Table B5. Final GSQ Ratings by Program Characteristics

			1-star	2-star	3-star	4-star	5-star
		Mean					
	Ν	(SD)	(%)	(%)	(%)	(%)	(%)
Program Type							
Licensed Child Care Center	2,106	3.74	1	31	648	1,258	168
		(0.63)	(0%)	(1%)	(31%)	(60%)	(8%)
Licensed Group Child Care Home	2,106	3.74	1	31	648	1,258	168
		(0.63)	(0%)	(1%)	(31%)	(60%)	(8%)
Registered Family Child Care Home	689	2.86	34	118	479	28	30
		(0.75)	(5%)	(17%)	(70%)	(4%)	(4%)
Alternative Pathways							
GSRP	1,086	4.05	1	4	74	872	135
		(0.46)	(0%)	(0%)	(7%)	(80%)	(13%)
Head Start	481	4.05	0	2	13	426	40
		(0.35)	(0%)	(0%)	(3%)	(89%)	(8%)
Early Head Start	163	3.92	0	2	16	138	7
		(0.43)	(0%)	(1%)	(10%)	(85%)	(4%)
Educational Approach							
Montessori	73	3.23	0	7	44	20	2
		(0.66)	(0%)	(10%)	(60%)	(27%)	(3%)
Reggio	68	3.46	1	5	31	24	7
		(0.84)	(2%)	(7%)	(46%)	(35%)	(10%)
Religious	294	3.23	3	20	191	66	14
		(0.69)	(1%)	(7%)	(65%)	(22%)	(5%)
Accreditation Status							
NAEYC	136	3.94	0	0	21	102	13
		(0.50)	(0%)	(0%)	(15%)	(75%)	(10%)
Non-accredited	1,970	3.73	1	30	562	1002	154
		(0.62)	(0%)	(2%)	(32%)	(57%)	(9%)
NAFCC	49	3.29	0	6	31	4	8
		(0.89)	(0%)	(12%)	(63%)	(8%)	(16%)

			1-star	2-star	3-star	4-star	5-star
		Mean	n (9/)	<i>n</i>	<i>n</i>	n (%)	<i>n</i>
Non accredited	N 1 641	(SD)	(%)	(%)	(%)	(%) E2	(%)
Non-accieulteu	1,041	2.07	120	411	1014	5Z	38 (20/)
Ages Conved Infonts and Taddlers		(0.76)	(7%)	(25%)	(62%)	(3%)	(2%)
Ages Served—Infants and Toddlers	2.547	2.05	400	400	4.470		
Yes	2,517	2.95	123	433	1478	415	68
		(0.80)	(5%)	(17%)	(59%)	(16%)	(3%)
No	1,273	3.92	3	14	212	898	146
		(0.58)	(0%)	(1%)	(16%)	(70%)	(13%)
Ages Served—Preschool							
Yes	2,195	3.56	19	92	889	1030	165
		(0.73)	(1%)	(4%)	(41%)	(47%)	(7%)
No	1,601	2.88	108	356	804	284	49
		(0.88)	(7%)	(22%)	(50%)	(18%)	(3%)
Locale							
Rural	1,228	3.28	59	151	486	451	81
		(0.93)	(5%)	(12%)	(39%)	(37%)	(7%)
Urban	2,568	3.27	68	297	1,207	863	133
		(0.83)	(3%)	(12%)	(48%)	(31%)	(6%)
Size							
Capacity of 0–34	525	3.86	1	11	116	332	65
		(0.66)	(0%)	(2%)	(22%)	(63%)	(13%)
Capacity of 35–58	535	3.73	0	4	176	316	39
		(0.60)	(0%)	(1%)	(33%)	(59%)	(7%)
Capacity of 59–100	605	3.69	0	10	214	336	45
		(0.63)	(0%)	(2%)	(35%)	(56%)	(7%)
Capacity of 101+	441	3.69	0	6	142	274	19
		(0.57)	(0%)	(1%)	(31%)	(60%)	(8%)

Note. GSQ is Great Start to Quality. SD is standard deviation. GSRP is Great Start to Readiness Program. NAEYC is National Association for the Education of Young Children. NAFCC is National Association for Family Child Care