



PATHS TO QUALITY RECOGNITION

How Child Care Programs Navigate Self-Reflection, Validation, and Observation in Michigan's Great Start to Quality

Rebecca Frausel, Neda Senehi, and MaryBeth Talbot

Introduction

Michigan launched the reimagined version of its Quality Recognition Improvement System (QRIS), Great Start to Quality (GSQ), in February 2023, following years of input-gathering and development.¹ The overall goal of the reimagined system is to ensure that all programs, both center-based and home-based, have opportunities to demonstrate and improve quality. Programs are encouraged to continuously strive for higher quality, with their progress recognized across five quality levels and supported by a system that promotes program choice.

QUALITY LEVEL PROGRESSION

All programs in good standing with licensing qualify for Level 1, Maintaining Health & Safety. Programs complete a Self-Reflection of their practices using 40 research-based program quality indicators to qualify for Level 2, Reflecting on Quality. After setting goals for quality improvement to qualify for Level 3, Enhancing Quality, programs apply for Validation and Observation. During Validation, evidence submitted to support proof of meeting the indicators is reviewed, after which programs qualify for Level 4, Enhancing Quality-Validated. To qualify for Level 5, Demonstrating Quality, programs must meet minimum threshold scores for each age



¹ Graber, et al. (2023). *Conditions for Change: Michigan's Great Start to Quality System*. Lansing, MI: Public Policy Associates.

https://publicpolicy.com/wp-content/uploads/2024/01/V4_PPA-Conditions-for-Change-1.pdf



group served on an observation tool that evaluates the program's environment, practices, and caregiver-child interactions.

PROMOTION OF PROGRAM CHOICE

Unlike the prior points-based system,² the reimagined GSQ allows programs more freedom in how to demonstrate quality. During Self-Reflection, programs respond to the indicators with one of the following options inclusive of all program types: (1) Currently meeting, (2) Not meeting at this time, (3) Not meeting at this time – Create a goal in the Quality Improvement Plan (QIP), and (4) Not aligned to program philosophy.³ To advance through the process, programs must be “Currently meeting” one indicator at minimum.

For Observation, programs serving infants and toddlers choose between the Environmental Rating Scales (ERS) and the Classroom Assessment Scoring System (CLASS). The ERS and CLASS have different versions for different ages, and the ERS has a version specific to home-based child care. The tools emphasize different aspects of care, with the ERS focusing on the holistic environment and the CLASS focusing on adult-child interactions. Programs serving only school-age children must use the Social & Emotional Learning Program Quality Assessment (SEL PQA), and GSRP classrooms must use the CLASS.

DIFFERENCES BY PROGRAM TYPE

Now in its third year of implementation, it is important to understand how different program types, especially center- and home-based programs, are responding to the reimagined system. Michigan has two types of licensed home-based programs, distinguished by the maximum number of children accommodated: group homes (12 children) and family homes (6 children). This brief examines how child care centers, group homes, and family homes engage with three core components of the GSQ: Self-Reflection, Validation, and Observation. Exploring differences by program types allows for a greater understanding of how the system is supporting quality advancement in different care settings, and whether further refinements are needed.

Study Overview: Public Policy Associates (PPA); the Michigan Department of Lifelong Education, Advancement, and Potential (MiLEAP); and the Early Childhood Investment Corporation (ECIC) are partnering for a four-year study of the implementation and impact of Michigan's reimagined Great Start to Quality (GSQ) system. Funded by MiLEAP and implemented by ECIC, GSQ aims to enhance the overall quality of child care in the state. The study examines the effectiveness of the new system in promoting program participation and quality level achievement by all programs. Methods include examining administrative data on GSQ participation, performance, and ratings, as well as collecting primary data on from programs, parents, and GSQ staff on experiences and satisfaction.

² Iruka, et al. (2018). *Great Start to Quality Program Validation Study Final Report*. Ypsilanti, MI: HighScope Educational Research Foundation. https://greatstarttoquality.org/wp-content/uploads/2024/11/GSQ-Validation_Final_Report_654524_7.pdf

³ For brevity, figures throughout this brief refer to these four options as: Meeting, Not Meeting, Goal, and Not Aligned.



Methods

The research team utilized Self-Reflection and Validation data from February 2023 through June 2025. During this period, 4,122 programs completed a Self-Reflection and 1,749 underwent Validation. The 40 quality indicators⁴ fall into 5 domains: Family and Community Partnerships (7 indicators); Inclusive Practices (4 indicators); Curriculum, Instruction, and Learning (15 indicators); Professional Development (9 indicators); and Staff Qualifications (5 indicators). Patterns of Self-Reflection responses were examined by program type and indicator domain. Among the programs that completed Validation, instances where Validation and Self-Reflection did not agree were identified.

Observation records from February 2023 to March 2025 were analyzed from 722 classrooms in 436 programs. The research team examined differences between different program types in observation tool selected and whether programs met the minimum thresholds for each age group/classroom. Each tool (CLASS, ERS, and SEL PQA) focuses on different quality areas (i.e., domains), and CLASS also has different domains based on the ages of children observed (see Table A3). The [Technical Appendix](#) provides additional details on the data sources and processes for Self-Reflection, Validation, and Observation, as well as more detailed results.

Results

Across the Self-Reflection, Validation, and Observation analyses, key differences between center-based and home-based programs emerged, particularly relating to how many indicators programs reported they are meeting, agreement between Self-Reflection and Validation, Observation tool choice, and pass rates.

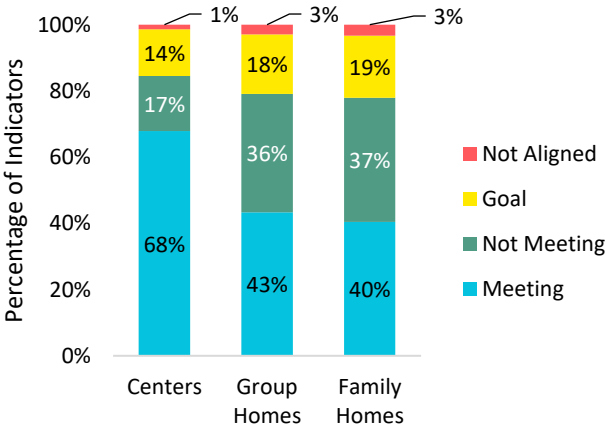
PROGRAMS’ SELF-REFLECTIONS

Centers reported they were currently meeting more indicators than group and family homes.

Licensing records show that between February 2023 and June 2025, 9,543 programs were licensed to operate in Michigan for at least one month (5,236 centers; 1,742 group homes; 2,565 family homes). Of centers who were open during this period, just over half (2,770, or 53%) completed Self-Reflections, compared to around a third (630, 36%) of group homes and relatively fewer family homes (722, 28%).

As shown in Figure 1, on average, centers reported on the Self-Reflection that they were currently meeting 68% of the indicators (that is, around 27 out of 40). This is

Figure 1. Average Percentage of Responses to Indicators by Program Type



⁴ For specific indicator text, see the guidance provided by ECIC as of July 2025, available here: <https://greatstarttoquality.org/wp-content/uploads/2025/08/GuidanceDocument7.2025English.pdf>

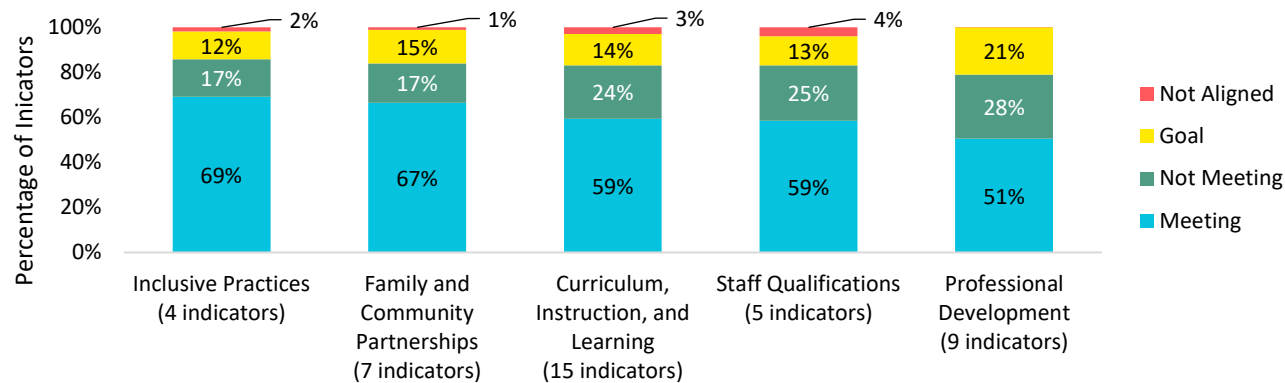


higher than group and family homes, which respectively averaged meeting 43% and 40% of the indicators (that is, around 16-17 out of 40). The number of indicators selected as goals was comparable across programs (14%-19%). Group and family homes (36%-37%) reported not currently meeting (and not setting as a goal) more indicators than centers (17%). Very few indicators were marked as not aligned with program philosophy (3% for group and family homes; 1% for centers).

Programs reported meeting fewer indicators in Professional Development, and often set these indicators as goals.

Among the indicators in the Professional Development (PD) domain, programs on average reported currently meeting only about half of them (51%). By contrast, programs reported currently meeting around 69% of the indicators in the Inclusive Practices domain (see Figure 2). The PD domain also had the highest prevalence of indicators programs intended to set as a goal in the Quality Improvement Plan (21%).

Figure 2. Average Percentage of Responses to Indicators by Indicator Domain



Among different program types, centers reported that they were currently meeting more indicators than group and family homes in each domain, mirroring the overall results from Figure 1. However, average rates of currently meeting the PD indicators were low even among centers, with center-based programs reporting they were currently meeting just 54% of these indicators on average (see Table A1).

VALIDATION OF SELF-REFLECTIONS

More center-based programs than home-based programs underwent Validation.

Of the programs that completed the Self-Reflection, 1,749 (42%) progressed to Validation during the time period studied. Just over half (51%) of centers underwent Validation, compared to 27% of group homes and 23% of family homes.

TABLE 1. NUMBER OF PROGRAMS WITH SELF-REFLECTION AND VALIDATION BY TYPE

PROGRAM TYPE	SELF-REFLECTION ONLY	SELF-REFLECTION WITH VALIDATION	TOTAL
Center	1,357 (49%)	1,413 (51%)	2,770 (100%)
Group Home	460 (73%)	170 (27%)	630 (100%)
Family Home	556 (77%)	166 (23%)	722 (100%)
Total	2,373 (58%)	1,749 (42%)	4,122 (100%)



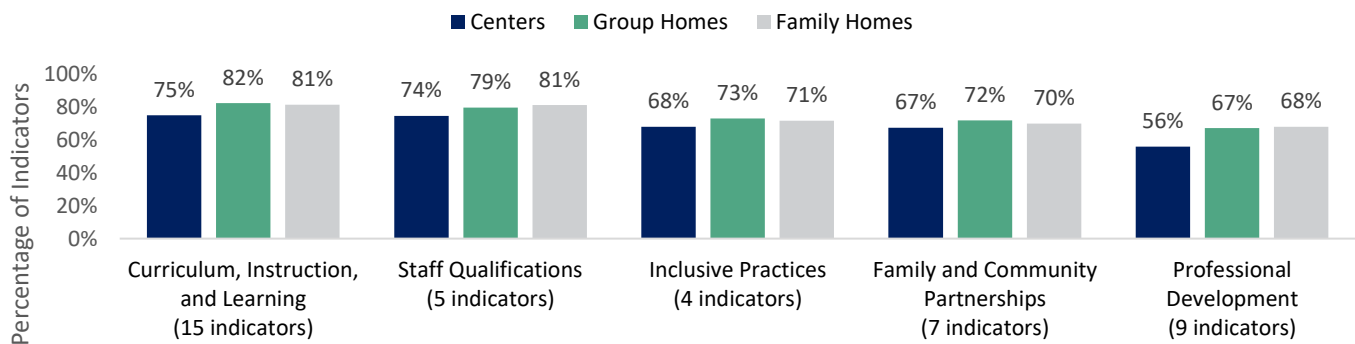
More than two-thirds of indicators on Self-Reflections with Validation agreed, with greater agreement among home-based programs.

Among the 1,749 programs that completed the Validation process of their Self-Reflections, the research team examined agreement between programs’ Self-Reflection and the GSQ Validators’ assessment of the evidence. Of the 66,889 indicators, 70% agreed (i.e., the GSQ Validator affirmed the programs’ self-assessment of the indicator, based on the evidence provided). The agreement rate was 75% for family homes and 76% for group homes, and slightly lower (68%) for centers. Most of the disagreements (90%) were instances where the program said, “Currently meeting” and the validator said, “Not currently meeting” (see Table A2 for more details).

Validators often disagreed with self-assessments of Professional Development indicators.

Across the different domains, group homes and family homes consistently showed higher rates of agreement between Self-Reflection and Validation indicator responses. All program types had lower rates of agreement for the Professional Development indicators (56% for centers, 67% for group homes, and 68% for family homes).

Figure 3. Average Percentage of Indicators with Agreement Between Self-Reflection and Validation by Domain



An indicator in the Family and Community Partnerships domain also emerged as a key area of disagreement between Self-Reflection and Validation: FCP2 (“Program has a family handbook or contract”). The agreement between Self-Reflection and Validation responses for this indicator was 27% for centers, 32% for group homes, and 36% for family homes.

CLASSROOM OBSERVATIONS

Most observations were completed in center-based classrooms.

Between February 2023 and March 2025, a total of 436 unique programs, completed observations, most of which were conducted in centers (372, 85%). The observation data includes far fewer group homes (35, 8%), and family homes (29, 7%). There were 18 center-based programs that used multiple tools (e.g., Toddler CLASS for a toddler classroom and SEL PQA for a school-age classroom), so in total, there were 454 unique program-tool observations.



CLASS was the most frequently used observation tool for all program types.

Across program-tool observations, CLASS was the most frequently selected, accounting for 85% of all program choices. In contrast, only 9% chose ERS and 6% chose SEL PQA. The ERS was slightly more common among group homes and family homes, while the SEL PQA was exclusively used by centers.

TABLE 2. OBSERVATIONS BY SELECTED TOOL AND PROGRAM TYPE

TOOL	OVERALL	CENTERS	GROUP HOMES	FAMILY HOMES
CLASS	384 (85%)	339 (87%)	25 (71%)	20 (69%)
ERS	42 (9%)	23 (6%)	10 (29%)	9 (31%)
SEL PQA	28 (6%)	28 (7%)	--	--
Total	454 (100%)	390 (100%)	35 (100%)	29 (100%)

Some centers also used different tool sub-types for different classroom ages (e.g., Toddler CLASS for a toddler classroom and Infant CLASS for an infant classroom); there were 639 unique program-tool subtype observations (see Table A4). Pre-K CLASS was the most widely used tool sub-type, accounting for 50% of all observations and was exclusively used in center-based programs.

Almost half of centers had multiple classrooms observed.

Among the 372 center-based programs observed, nearly half (163, 44%) had more than one classroom observed (often using different tools or tool sub-types for different classrooms). Group and family homes were always scored using one classroom and one tool. Center observation scores reflect the variability that exists across different classrooms and teachers, while scores for group and family homes reflect a single classroom environment.

Most programs with Observation achieved the minimum threshold scores.

Overall, about 70% of programs met the minimum threshold scores across all observed classrooms and tools, meaning they “passed” the observation (see Table 3, bottom left cells). However, pass rates varied depending on program type and tools used, as can also be seen in the table.

CLASS showed higher pass rates compared to ERS and SEL PQA.

Programs were more likely to pass the observation when observed using the CLASS. Around 77% of programs that completed their observation using CLASS passed, compared to 38% for programs using ERS and 21% for programs using SEL PQA (see Table 3, left Overall column). Pass rates between different program types are identical between centers (76%) and group homes (76%) when using CLASS, and family homes show a slight advantage (85%) (see Table 3, CLASS row).

At the classroom level, 82% of classrooms that were observed with CLASS met thresholds, with Infant classrooms performing best (89% pass), followed by Pre-K classrooms (84% pass) and Toddler classrooms (76% pass) (see Table A4 in the Technical Appendix for complete classroom-level results).



The ERS is challenging to pass, especially for family homes, but is also infrequently chosen.

As already noted, fewer programs met the minimum thresholds on the ERS (38%) and SEL PQA (21%), suggesting demonstrating quality may be more difficult using these tools. When using the ERS, 3 out of 9 family homes passed the observation, compared to 4 out of 10 group homes and 9 out of 23 centers.

The SEL PQA showed the lowest pass rates.

The SEL PQA was only used in a small number of center-based programs and had the lowest pass rate of all the observation tools; just 21% of programs met the minimum-threshold scores.

TABLE 3. PERCENTAGE OF PROGRAMS MEETING MINIMUM THRESHOLDS BY TOOL AND PROVIDER TYPE

PROGRAM TOOL	PASSED/DID NOT PASS	OVERALL	CENTERS	GROUP HOMES	FAMILY HOMES
CLASS	PASSED	295 (77%)	259 (76%)	19 (76%)	17 (85%)
	DID NOT PASS	89 (23%)	80 (24%)	6 (24%)	3 (15%)
	TOTAL	384 (100%)	339 (100%)	25 (100%)	20 (100%)
ERS	PASSED	16 (38%)	9 (39%)	4 (40%)	3 (33%)
	DID NOT PASS	26 (62%)	14 (61%)	6 (60%)	6 (67%)
	TOTAL	42 (100%)	23 (100%)	10 (100%)	9 (100%)
SEL PQA	PASSED	6 (21%)	6 (21%)	--	--
	DID NOT PASS	22 (79%)	22 (79%)	--	--
	TOTAL	28 (100%)	28 (100%)	--	--
TOTAL UNIQUE PROGRAM-TOOL OBSERVATIONS	PASSED	317 (70%)	274 (70%)	23 (66%)	20 (69%)
	DID NOT PASS	137 (30%)	116 (30%)	12 (34%)	9 (31%)
	TOTAL	454 (100%)	390 (100%)	35 (100%)	29 (100%)

Taken together, these findings suggest that some observation tools may be more accessible for programs to meet minimum quality thresholds than others. When using CLASS, most programs met the minimum-threshold scores, while fewer were able to do so with ERS and SEL PQA. These differences may not so much reflect differences in program quality but instead reflect differences in tool structure and how quality is defined and measured.

Discussion

Conclusions

This study provides an early look at how Self-Reflection, Validation, and Observation are functioning for different program types in Michigan's reimagined GSQ system. Key differences between center-based and home-based programs were identified that shed light on both system strengths and ongoing challenges.

During Self-Reflection, center-based programs were more likely to report currently meeting more quality indicators than home-based programs, consistent with the fact that larger programs may have greater administrative capacity and resources to document quality practices. However, both center- and home-based programs struggled with Professional Development indicators, both in Self-Reflection and Validation.



Validators frequently disagreed that programs' evidence supported meeting these indicators, suggesting that programs lack clarity on what constitutes sufficient evidence, or face barriers to staff training and advancement. This highlights the importance of targeted supports for workforce development, a persistent challenge in the child care sector.

Additionally, program type related to whether and how programs participated in Validation. Centers were more likely to undergo Validation, but their agreement rates between Self-Reflection and Validation were lower than the agreement rates of home-based programs. Centers may be submitting more optimistic assessments, or home-based programs may be more conservative and cautious when reporting quality practices. Program types differ in how they navigate the Self-Reflection and Validation components of GSQ, highlighting opportunities for tailored technical assistance that focuses on reach for home-based programs and fidelity for center-based programs.

Analyses of the observations reveal important differences in how center- and home-based programs fare when using different tools. CLASS was the most widely used and had the highest pass rate across all program types. By contrast, pass rates were lower with the SEL PQA (required for school-age classrooms or programs) and the ERS. Centers face a higher bar since they have multiple classrooms and all classrooms must meet the minimum threshold so one low-scoring classroom can prevent a center-based program's advancement. Additionally, all programs using CLASS face a higher bar since all their classrooms must meet the minimum threshold for all CLASS domains in a sub-tool, so classrooms with one low-scoring domain can prevent a program's advancement to Level 5. Despite these dual challenges, center-based programs and programs choosing CLASS have higher pass rates.

Although the ERS tools were originally intended to put home-based programs on a more even footing with centers, the observation data to this point suggests that very few home-based programs selected ERS and among those that did, pass rates were notably lower than for CLASS. The ERS tools may be harder to pass relative to CLASS, or programs may have less experience with ERS. Additionally, there may be selection bias, such that programs that opt into using ERS may be systematically different than those choosing CLASS (e.g., newer programs, less resourced, or those struggling with quality already), which may contribute to lower pass rates. It is important to note that very few programs selected ERS, which limits the generalizability of these findings.

Overall, these results highlight a potential disadvantage for home-based programs that use ERS tools. This finding points to questions for both research and policy to consider whether thresholds should be recalibrated for home-based settings, whether targeted training and support are needed, or whether alternative tools may better capture quality in these care settings. Understanding whether the observed pattern reflects limitations in tool, provider characteristics, or contextual factors will be important for guiding state policy decision about observation tools and ensuring equity across provider types.

Suggestions for Policy

These findings point to several areas where refinements to GSQ could strengthen ECIC's ability to support quality improvement across different care settings:



- Promote the participation of home-based providers throughout the Self-Reflection, Validation, and Observation processes, and understand specific barriers and challenges home-based providers face with these aspects of GSQ participation.
- Clarify evidence expectations for programs seeking Validation to increase agreement with Self-Reflection responses, particularly for Professional Development (PD) indicators.
- Explore ways to strengthen the relationship and data pipeline between ECIC and MiRegistry (Michigan's professional development registry for the early childhood and out-of-school time workforce). For example, advertise which PD indicators specific trainings apply to, or use training completion records to auto-populate quality indicator evidence.
- Provide additional technical support and training for programs choosing to use the ERS, or who are required to use the SEL PQA because they serve school-age children. The consistently lower selection and pass rates for these tools may reflect these tools' resource- and training- intensive requirements, as opposed to practical differences in program quality.

Next Steps

As the study enters its fourth and final year, the research partnership is focused on ensuring a comprehensive picture emerges of how changes to Michigan's GSQ impacts program practices and families' access to high-quality child care. Next steps broadly include delving deeper into specific data sources, as well as combining evidence across different sources. More specifically:

- Track participation patterns and quality-level attainment of center- and home-based programs to understand how different programs are navigating the system longer term.
- Combine administrative data on program quality-level achievement with MiRegistry employment records to examine the mediating factor of staff turnover on quality improvement.
- Examine whether programs that receive specific trainings or technical assistance to prepare for Observation are more likely to pass the threshold scores for their chosen tool.
- Compare validated quality indicators with observation scores to understand which indicators most strongly predict these unbiased observational measures of program practices.
- Link primary data gathered from child care programs, families, and GSQ staff to compare how different perspectives and experiences shape trajectories. For example, providers who perceive the GSQ system as more aligned with their views of quality may be more likely to progress in levels.