



NEW JERSEY DEPARTMENT OF
CHILDREN AND FAMILIES

Monitoring Methodology Summary

Office of Applied Research and Evaluation
Office of Monitoring

Version 1.6
July 2, 2025

Contents

Tool Development.....	3
Initial Development.....	3
Document Review	3
Initial Tool Drafting.....	3
Content Validity	4
Usability Testing	4
Interrater Reliability	4
Monitoring Process.....	4
Sample Selection	4
Participant Recruitment.....	5
Case Review Data Access and Collection.....	5
Data Recording.....	5
Quality Control	5
Scoring.....	5
References	6
Appendix A –FPS Monitoring Details	9
Sample Selection.....	9
Participant Recruitment.....	9
Case Review Data Access and Collection	9
Appendix B –KFT Monitoring Details	9
Sample Selection.....	9
Participant Recruitment.....	9
Case Review Data Access and Collection	9
Appendix C– DVDS Monitoring Details	9
Sample Selection.....	9
Participant Recruitment.....	10
Case Review Data Access and Collection	10
Appendix D – CMO Monitoring Details.....	10
Sample Selection.....	10
Participant Recruitment.....	10
Case Review Data Access and Collection	10

Monitoring Methodology

The Department of Children and Families (DCF) launched the Office of Monitoring (OOM) as a key strategy for delivering high-quality services to New Jersey (NJ) constituents, a priority outlined in DCF's strategic plan. OOM aimed to improve systematic monitoring of DCF's contracted services by ensuring clear performance standards, building and implementing monitoring methods and tools, and creating and maintaining consistent feedback loops with key stakeholders (NJ DCF, 2021). This restructuring allowed DCF to assess the aspects of services that constituents value most (NJ DCF, 2021). The process was guided by listening to participant needs and ensuring those with lived experience are included in decisions.

DCF selected the Availability, Accessibility, Acceptability and Quality (AAAQ) framework (Jensen et al., 2014) to monitor programs based on a literature review, feedback gathered from People with Lived Experience (PWLE), and input from various stakeholders during the Commissioner's listening tour and public comment periods.

OOM partnered with the DCF Analytics and Systems Improvement (ASI) to tailor the AAAQ framework for DCF's contracted services and to develop monitoring processes by building data collection tools that are summarized into provider-specific Program Performance Scorecards. Below is a description of the process DCF used to develop tools for each service line.

Tool Development

Initial Development

DCF implemented a multi-phase process to build a framework, performance standards, data collection tools and methods for monitoring quality of service delivery among contracted provider agencies. To identify an organizing framework for the quality domains of primary concern to stakeholders, DCF first gathered information from stakeholders themselves. ASI conducted a qualitative meta-synthesis of all known needs assessments (N=9) carried out by or for the Department between 2012 and 2019, representing the perspectives of parents, youth, foster parents, DCF staff, provider agencies and community partners. They extracted key themes, then coded them using an inductive approach. From this analysis arose an organizing framework consisting of cross-cutting concepts related to high quality service delivery. Concepts mapped onto the domains of the United Nations' AAAQ framework, which highlights four human rights-based standards for service provision: availability, accessibility, acceptability, and quality of services (Jensen et al., 2014). To ensure this framework adequately captured constituents' voices, DCF then conducted a series of focus group discussions and interviews with parents, youth previously involved with NJ's Child Protective Services, DCF staff and provider organizations to vet the AAAQ for relevance and identify any gaps in the framework.

Document Review

The tool development process began by meeting with program administrators and program leads to collect programmatic documents such as program-specific RFPs, contracts, logic models, practice profiles and program manuals. The DCF tool development team reviewed the documentation prior to a second meeting with program staff where the team discussed any remaining questions about the program. DCF cross-walked the AAAQ Framework with contract documents and requirements/standards from relevant accreditation organizations to identify standards that appeared in the contracts across agencies and any possible overlap with accreditation standards.

Initial Tool Drafting

To operationalize the AAAQ into a set of performance standards, data collection tools and methods, DCF applied Lynn's two-stage content validation process (Lynn, 1986) which included a developmental stage and a quantification stage. In the developmental stage, DCF reviewed service-specific documents and noted measurable objectives. They also conducted a literature review to identify any existing measures of concepts mentioned in the AAAQ framework and program documentation. The team used this information to develop a set of cross-service and service-specific items for case record reviews and service participant interviews (or surveys for Domestic Violence, DV).

Content Validity

During the quantification stage, ASI then convened an expert workgroup consisting of OOM, DCF program staff and representatives from each of the providers to review the draft tools and obtain any necessary additional information. ASI provided the expert workgroup with a spreadsheet that contained each item developed for the tools and columns to rate how relevant and clear each item was on a scale of 1 not relevant/clear to 4 highly relevant/clear, as well as an area to write additional comments.

ASI then calculated the average ratings of relevance and clarity for each question and interrater agreement scores (the extent to which work group members' ratings agreed with each other) using the data, then facilitated a focus group to obtain additional qualitative feedback. For example, if an expert rated a question with low clarity, the facilitator asked them for suggestions to improve the question. If there was disagreement about relevance, with some high and some low ratings, ASI asked the group to discuss their ratings for that question to arrive at a consensus. DCF used this data from the workgroup of experts to select the most relevant questions for the tools.

Usability Testing

The next step in DCF's tool development process was to assess whether (1) the tools were user-friendly, and (2) how important each of the AAAQ standards were to measure program quality from the perspectives of PWLE. To test the participant interview (or survey for DV) and case record review tools, DCF monitors conducted practice case reviews and interviews (or surveys for DV) with participants who received program services. DCF then gathered feedback about tool usability in three focus groups: one with PWLE focused on the interview tool*; one with monitors focused on the interview tool; and one with monitors focused on the case record review tool. In addition, after experiencing a practice interview, PWLE were asked to rate the importance of each standard and discuss whether DCF should consider any additional measures. The feedback was used to create final drafts of the tools.

*Note: DCF did not hold a focus group for DV to protect participant confidentiality; this information was instead collected as part of the survey.

Interrater Reliability

DCF then assessed the extent to which the DCF monitors agreed with each other when completing the case record review tool to ensure consistency across raters as well as clarity of items and tool completion guidelines. Each monitor used the case record review tool and guidelines to review and rate three cases. DCF used the ratings to calculate Fleiss' kappa for the overall tool, and the Intra-Class Correlation to examine consistency among raters. DCF also calculated the percent agreement among monitors for each individual question. DCF discussed any questions where 3 or more raters disagreed in a meeting to determine if monitors needed further training, if the tool needed to be revised, and if the guidelines needed to be clarified.

Monitoring Process

Note: The below Sample Selection and Participant Recruitment sections do not apply to DV. See Appendix C for information on sample selection and participant recruitment for DV programs.

Sample Selection

DCF selected cases from a pre-defined period prior to monitoring, ensuring that providers had sufficient time for data entry and submission. For the majority of programs, DCF used a standard z-score of 1.96 with a 95% confidence interval and a 10% margin of error to calculate the total number of cases to be monitored. These calculations took the size of each site into account.

If a given site had less than 12 total cases, the team included all cases in the random sample. If a site had more than 12 total cases, DCF provided 20% of the random sample size for each stratum to be utilized as replacements if a participant was ineligible, unreachable or refused to participate. In these instances, DCF continued to use the original sample for case record reviews. DCF attempted to obtain the highest response rate possible but acknowledged that there would be a smaller number of participant interviews than case record reviews.

At this stage, DCF did not include families whose primary language was not English or Spanish in the monitoring process; however, in the future the Department intends to translate materials into more languages.

Participant Recruitment

Following sample selection, DCF obtained demographic and contact information for the participants in the sample for the purpose of recruitment. This information included phone number and email address when available (see Appendix A for more detailed information). If an email address was available, monitors began recruiting the participants with an email in the morning that explained the purpose of the interview and let the participant know they would receive a call from the monitor later that day. That afternoon the monitor sent a text to the participant (or call if it was a land line) inviting them to participate. The next day they called the participant in the morning, then made a final call in the afternoon. If a participant refused or agreed to participate at any step, monitors did not conduct the subsequent contact steps. The monitors offered to do the interview immediately when they first spoke to the participant but may also have scheduled it within the following few days if the participant was not available at that time. After the final call, the monitor stopped contacting the participant if they had not received a response.

Case Review Data Access and Collection

DCF notified agencies approximately one month prior to when the monitoring process began. Agencies provided DCF with access to (redacted, if appropriate) case record files for monitors to read cases and rate them for each item in the case review tool.

Data Recording

Monitors recorded the participant interviews and case record reviews on paper, and subsequently entered all data in REDCap, a secure, web-based application designed to support data capture for research studies. When monitors were able to complete both a participant interview and case record review for a single participant, a unique identifier linked the two tools.

Quality Control

The Office of Monitoring performed a quality control (QC) process on 25% of the sample of all cases reviewed per unit. This process consisted of a simultaneous second review of a case(s) by the Unit Supervisor, where they read the whole case then compared their results to the monitor's tool. The Unit Supervisor discussed any noted discrepancies with the monitor and reconciled them, involving OOM leadership as needed in situations where inconsistencies between the Unit Supervisor and the monitor required further support and guidance.

Supervisors also performed a full spot check to assess safety and ensure data reliability for 50% of the remaining sample that did not receive a second review under the quality control process. For example, if the total sample was 100 cases, 25 received a complete Quality Control review, then 38 cases (50% of the remaining 75) received a full spot check. Full spot checks involved using a standardized tool to assess unaddressed allegations, threats to safety and risk, service needs or interventions, and inconsistencies across items without reading the entire case. If a full spot check yielded less than 90% correct or consistent responses, all cases assigned to that monitor/trainee were subject to a full spot check review.

When the quality control process was complete, ASI also conducted an independent review of the data to clean any remaining quality issues.

Scoring

After data cleaning, ASI used the mean of each individual item to calculate the mean standard score across tools (see Figure 1). ASI then repeated this process to calculate the overall domain scores (see Figure 2). ASI produced a scorecard using Tableau that indicates the average scores for individual items, standards and domains by site. The scorecards show all data for case record reviews. However, if any interview item had less than three responses, it was excluded from the scoring to protect participant confidentiality. There is also a column that displays the average for all sites statewide to facilitate comparison of how the individual site performed compared to the entire network of providers for that service line. DCF used scores for the first round for informational purposes, to familiarize providers with the scorecard, inform further tool development (if needed) and engage in collaborative quality

improvement efforts.. After the initial pilot, DCF will use the scorecards for a formal collaborative quality improvement process. This process, known as CoQI, is a collaborative and structured approach focused on systematically evaluating and improving practices, policies, and procedures through data-driven decision-making. CoQI involves a series of meetings featuring structured activities to analyze program performance, identify root causes of challenges, and develop and implement effective improvement strategies.

Figure 1. Averaging of Standard Scores

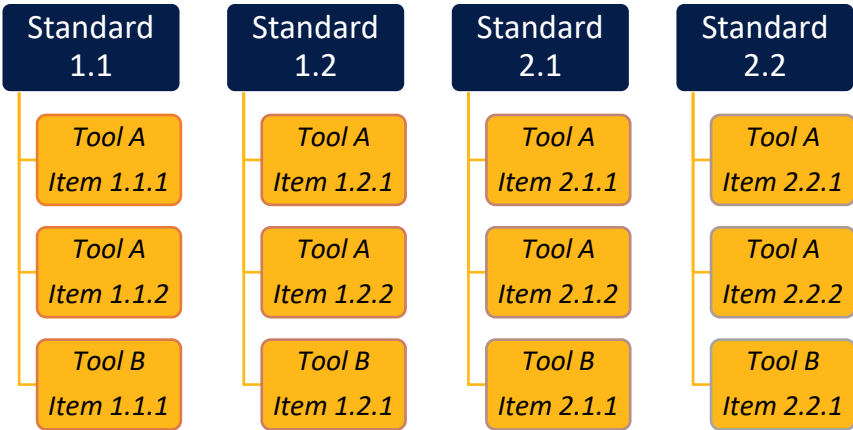
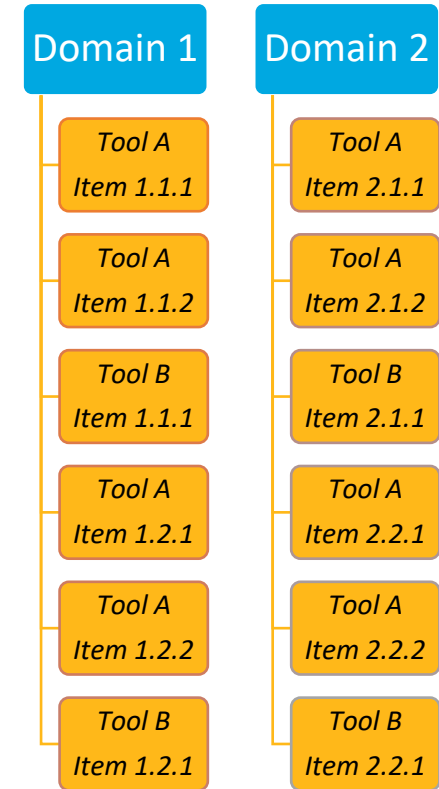


Figure 2. Averaging of Domain Scores



References

Jensen, M. H., Villumsen, M., & Petersen, T. D. (2014). *The AAAQ framework and the right to water: International indicators for availability, accessibility, acceptability and quality: An issue paper for the AAAQ toolbox*. Danish Institute for Human Rights.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing research*, 35(6), 382-386.

New Jersey Department of Children and Families (2021). *Quality service, strong outcomes: Improving monitoring of DCF's contracted services*. Retrieved from www.nj.gov/dcf/providers/contracting/DCF-monitoring-concept-paper-7.7.2021.pdf.

Monitoring Appendices

Office of Applied Research and Evaluation
Office of Monitoring

Version 0.2
July 2, 2025

Appendix A –FPS Monitoring Details

Sample Selection

Cases were selected from a six-month period ending three months prior to when monitoring began.

DCF monitored completed cases and turnbacks, but not interrupted interventions. Completed cases were selected if their discharge date fell within the period of review, and turnbacks were selected if their referral date falls within this period. DCF removed families who were not turnbacks and not yet discharged at the time of data collection from the sample. Families who enrolled twice during the report period were only counted once, with their most recent case selected for review. If a family was a turnback and also completed services at another time during the period of review, they were considered eligible to be reviewed as a completed case.

Site is equivalent to county. In other words, if an agency serves multiple counties, each county is considered a separate site. Groups are defined by services provided and turnback categorization.

Participant Recruitment

DCF obtained recent contact information from NJ's child welfare administrative data system. Monitors also conducted the case record review prior to recruitment to confirm if the case records contained any additional contact information.

Case Review Data Access and Collection

Agencies provided DCF with access to web-based data systems when available or pulled cases when not available, and provided a space on-site for monitors to read cases and rate them for each item in the case review tool.

Appendix B –KFT Monitoring Details

Sample Selection

Cases were selected from a twelve-month period ending two months prior to when monitoring began.

DCF monitored cases who are not yet housed, currently housed, or discharged. Discharges were selected if their discharge date falls within the report period. Currently housed participants were selected if they had not been discharged or were discharged after the report period, and if their move-in date was at least one year prior to the start of the report period. Not yet housed participants were selected if they did not have either a discharge or move-in date. DCF removed cases who were screened eligible but did not make it to the provider, as well as cases that made it to the provider but were deemed ineligible.

Site is equivalent to office. In other words, if an agency has multiple offices, each office is considered a separate site. Groups are defined by housing status.

Participant Recruitment

DCF obtained the list of participants served during the report period from the KFT Master Family File provided to DCF by providers on a quarterly basis. DCF then selected the sample from this pool and requested contact information for these cases from the agencies by whom they were served.

Case Review Data Access and Collection

Agencies provided DCF with access to web-based data systems when available or paper case files if they were not electronic, and provided a space on-site for monitors to read cases and rate them for each item in the case review tool.

Appendix C– DVDS Monitoring Details

Sample Selection

Cases were selected from a six-month period ending two months prior to when monitoring began.

DCF monitored both residential and non-residential cases. Cases were selected if their discharge date fell within the report period and were not currently receiving services. Residential cases are defined as those that utilized emergency shelter while non-residential cases did not utilize emergency shelter. DCF requested providers send a list of adult cases who were discharged during the six-month report period, and then selected the sample from that list. DCF did not monitor cases that only contacted the hotline, as these do not have a case record. Cases who were discharged more than once during the report period were only counted once, with their most recent case selected for review. DV cases were redacted due to confidentiality laws and provided electronically.

Site is equivalent to agency. In other words, where a county is served by multiple agencies, each agency is considered a separate site. Groups are defined as case type (residential or non-residential).

Participant Recruitment

Due to the sensitive nature of DV data, DCF is not allowed to collect any identifying information for participants. Therefore, this program utilizes a both a paper and web-based participant survey rather than a one-on-one interview. DCF asked providers to provide a survey to all participants who were discharged during the report period. DCF asked providers to give a paper survey to participants who were discharged in person, with an envelope they could seal before giving it back to the provider to ensure confidentiality and a better response rate. DCF monitors collected the paper surveys after the review period was complete and entered the data into a web-based system called REDCap, a secure, web-based application designed to support data capture for research studies. For participants who were discharged remotely, DCF asked providers to email or text them a direct link to the web-based survey.

Case Review Data Access and Collection

Agencies supplied DCF with a list of all case IDs meeting the inclusion criteria as specified above, from which DCF selected a random sample and provided the list of resulting case IDs to the agencies. Agencies then provided DCF with redacted case records for monitors to read cases and rate them for each item in the case review tool. Monitors recorded the case record reviews on paper and then entered all data in REDCap. The participant survey and case record review data are not linked in any way to ensure confidentiality.

Appendix D – CMO Monitoring Details

Sample Selection

DCF selected cases from a report period spanning from September 1, 2023, to February 29, 2024. DCF monitored closed cases of youth up to age 18 who had an episode of CMO care for at least four months during the report period. At this stage, DCF excluded from the monitoring process families whose primary language was neither English nor Spanish. In addition, cases where parental rights were terminated, and the local office manager was assigned as the guardian were excluded from the sample.

For CMO, site is equivalent to each agency's individual location. In other words, each agency is considered a separate site. Cases were randomly selected from each site, following the above criteria.

Participant Recruitment

DCF obtained recent contact information from the Child & Youth Behavioral Electronic Record database (CYBER), which is DCF's data system for the Children's System of Care.

Case Review Data Access and Collection

Monitors accessed the CMO case records in CYBER to review cases and rate each item using the case review tool. Monitors recorded participant interviews and case record reviews on paper and subsequently entered all data into REDCap, a secure, web-based application designed to support data capture for research studies.