

Grantee Performance Management System (GPMS) Draft Testing Plan

V5

GPMS Validation Team

- Alberto Sorongon, PhD
- Rick Fiene, PhD
- Thom Flottesmesch, PhD
- Traci Padgett, PhD

EAS Development Team

- Marisa Russo, PhD
- Melissa Bandy

Table of Contents

.....	1
1. STATUS UPDATE.....	3
2. GOALS OF TESTING PLAN.....	3
3. STAGE 1: FEASIBILITY TEST.....	3
4. STAGE 2: PILOT TEST.....	5
PILOT TEST SAMPLING AND RECRUITMENT.....	6
PILOT TEST METHODOLOGY.....	8
APPENDIX A: STAGE 2 PILOT TEST SAMPLING APPROACH.....	12

Grantee Performance Management System (GPMS) Draft Testing Plan V5

1. Status Update

As of June 14, 2021, the EAS Development Team and Validation Team are in internal meetings as well as discussions with OHS on the testing procedures outlined in this plan. The plan indicates where the team is still exploring testing components. Some components that are still under review include:

- Staffing plan
- Sample size and sampling approach
- Recruitment/incentive plan
- Testing design components (e.g., mode test for FA2)

This draft testing plan has been reviewed by members of the EAS Development Team and the Validation Team. A summary of the current testing plan was shared with OHS on 5/19/2021. An updated version was sent to DLH on 6/14/2021.

2. Goals of Testing Plan

The testing plan was developed with three (3) general goals in mind:

Goal #1: Refine FA1 and FA2 guides and EAS to effectively collect performance data to support informative EAS and GPMS scores

Data from both the Stage 1 Feasibility Test and the Stage 2 Pilot Test will address Goal #1.

Goal #2: Ensure that staff are enabled to implement the FA1 and FA2 guides and EAS in IT-AMS, supported by effective training and other necessary supports

Data from the Stage 2 Pilot Test will address Goal #2.

Goal #3: Finalize the EAS and GPMS scoring algorithms to reflect meaningful performance variance in reviewed grantees

Data from the Stage 2 Pilot Test will address Goal #3.

3. Stage 1: Feasibility Test

The primary focus of the Stage 1 Feasibility Test will be to test the flow and wording of questions in the data collection guides in order to refine the language of the FA1 and FA2 guide items, performance measures (PMs), quality indicators (QIs), and quality marker (QM) ratings before programming in IT-AMS for the Stage 2 Pilot Test. Note that the Feasibility Test will be conducted outside of IT-AMS.

The Feasibility test will be conducted virtually with convenient samples of 1-2 grantees for FA1 and 1-2 grantees for FA2. The Feasibility Test staff will be comprised of Lewin and DLH project staff, DLH review leads and reviewers, and OHS central office staff. For each review type, the Feasibility Test staff will conduct the test by reading the protocol guide items to the grantee participant who will answer the items accordingly. The Feasibility Test staff will record the responses according to the protocol guide instructions and use those responses to complete the QM ratings.

Information on the data collection process will be collected from both the respondent and the Feasibility Test staff conducting the Feasibility Test review:

- *Respondent Experience Data:* After each section, reviewer will ask respondent if there were any questions that the respondent did not understand or needed clarification and why and record their responses.
- *Feasibility Test Staff Experience Data:* The Feasibility Test staff will record information on the work flow, item wording, rating anchors, and review materials (including manual) (e.g., what did not work or was unclear and why) through a questionnaire/notes/diary. Feasibility Test staff will also record timing estimates for implementing the guides and completing the ratings.

The EAS development team will use the collected information to refine work flow, item wording, rating anchors, and review materials (including the review manuals) prior to finalizing technical/business requirements for IT-AMS programming. Notes and scoring data will also be collected (as feasible) to inform scoring refinement.

After revising the protocols based on the information collected from the Stage 1 Feasibility Tests, the EAS development team will meet with OHS to walk through the team's recommendations for the final protocols, guides, and EAS to describe the features and answer questions. Once OHS approves the review materials, the EAS development team and DLH' technical team will complete the technical and business requirements for programming into IT-AMS in preparation for the Stage 2 Pilot Test. See **Table 1** for an overview of the Stage 1 Feasibility Test schedule.¹

¹ Note that these dates are based on assumptions as of June 14, 2021. The DLH team will continually review this timeline throughout the development and testing period to report status of project activities, identify and mitigate risk to the schedule, and develop alternatives for OHS to consider if events impact the schedule. Note that these dates are continually being compared to the overall project timeline.

Table 1: Stage 1 Feasibility Test Schedule

Task Description	Completion Date(s)
FA1	
Draft FA1 protocol, data collection guides (interviews), EAS (PMs, QIs, QMs) ready for feasibility test	8/13/21
Conduct feasibility test of FA1 protocol, data collection guides (interviews, etc), EAS (PMs, QIs, QMs)	8/16/21-8/27/21
Submit draft FA1 protocol, data collection guides (interviews), EAS (PMs, QIs, QMs) based on feasibility test data to OHS for review	9/20/21
Meetings with OHS to walk through FA1 protocol, data collection guides, EAS (PMs, QIs, QMs) and answer questions	9/20/21-10/01/21
Approval from OHS on FA1 protocol, data collection guides (interviews), EAS (PMs, QIs, QMs)	10/04/21
FA2	
Draft FA2 protocol, data collection guides (interviews; explorations; file reviews; etc), EAS (PMs, QIs, QMs) ready for feasibility test	7/16/21
Conduct feasibility test of FA2 protocol, data collection guides (interviews, etc), EAS (PMs, QIs, QMs)	7/19/21-7/30/21
Submit draft FA2 protocol, data collection guides (interviews, etc), EAS (PMs, QIs, QMs) based on feasibility test data to OHS for review	8/30/21
Meetings with OHS to walk through FA2 protocol, data collection guides, EAS (PMs, QIs, QMs) and answer questions	8/31/21-9/10/21
Approval from OHS on FA2 protocol, data collection guides (interviews, etc), EAS (PMs, QIs, QMs)	9/13/21

4. Stage 2: Pilot Test

Data from larger Stage 2 Pilot Test will address all three of the goals of the GPMS testing plan:

- Goal #1: Refine FA1 and FA2 guides and EAS to effectively collect performance data to support informative EAS and GPMS scores
 - Survey data on the review experience of the pilot test respondents and staff (similar to those described in the feasibility test) will inform refinement of the guides and EAS items. Data from both the Stage 1 Feasibility Test and the Stage 2 Pilot Test will address Goal #1.
- Goal #2: Ensure that staff are enabled to implement the FA1 and FA2 guides and EAS in IT-AMS, supported by effective training and other necessary supports
 - Survey data on the effectiveness of the training sessions and the review support material (e.g., reviewer manual) will inform refinement of the training programs and support materials. Project design staff will also collaborate with DLH field management and training staff, as well as OHS to refine the reviewer and review lead requirements to ensure staff qualifications are appropriate for their assigned tasks.

- Goal #3: Finalize the EAS and GPMS scoring algorithms to reflect meaningful performance variance in reviewed grantees
 - EAS data from the QM ratings will be used to test the draft algorithms that were developed based on historic monitoring data as well as simulated data. The GPMS validation team are also exploring incorporating an interrater reliability component into the pilot test design.

Pilot Test Sampling and Recruitment

Pilot Test Sample

The Pilot Test will be conducted with a purposive sample of approximately 45 grantees for FA1 and 45 grantees for FA2 (see **Appendix A** for power analysis to determine recommended Pilot Test sample sizes).

A primary consideration of the purposive sampling approach is achieving sufficient variance in the QM ratings, which is critical to the success of this pilot test. As a result, the approach for pilot test sample selection of grantees will primarily focus on projected grantee performance levels (low-medium-high). Data to identify grantees projected to be low-, medium-, and high-performers will include analysis of their past monitoring results; recommendations by OHS, regional offices, and monitoring review staff; and analysis of preliminary risk indicator data. Other program characteristics (e.g., region, program option, program size) will also be considered in selecting pilot test grantees. Note that OHS has determined that they do not want the pilot test sample to include any grantees that are included in the FY2022 national monitoring schedule. As a result, grantees that are scheduled for a FY2022 monitoring review will not be eligible for the pilot test.

The Validation Team will collaborate with DLH project staff and OHS to analyze monitoring data and other collected information on the pilot test-eligible grantees and generate a list of grantees to recruit for the pilot test.

As mentioned above, the target sample size for the pilot test is 45 grantees for the FA1 pilot test and 45 grantees for the FA2 pilot test. We anticipate a 25 percent grantee refusal rate. To account for the grantees that will refuse an invitation, the Validation Team will create recruitment samples of 56 grantees for FA1 pilot and 56 grantees for the FA2 pilot.

Pilot Test Grantee Recruitment

While we have planned for a projected 25 percent refusal rate, pilot test sample recruitment activities will focus on activities to fully inform grantees on the activities and expectations for their participation in the pilot test, answer any questions or address any concerns grantees may have, and share the benefits (including incentives) for the grantees should they decide to participate.

The Pilot Test Recruitment staff will be comprised of DLH staff with experience communicating with grantees and scheduling monitoring reviews. Prior to contacting grantees in the recruitment sample, the Pilot Test Recruitment staff will be trained on the

recruitment approach, the pilot test goals and methodology, expectations for participating grantees, and answers to frequently asked questions (FAQs).

Once the recruitment sample is developed, Pilot Test Recruitment staff will email sampled grantees informing them that they have been selected for an invitation to participate in a pilot test of the new monitoring system tools and procedures. The email will inform the grantee that participation is voluntary, and will request dates and times for a phone call to discuss the pilot test. Once scheduled, the Pilot Test Recruitment staff will send an invitation to the grantee with contact information for the Pilot Test Recruitment call.

As part of the messaging plan for the Pilot Test Recruitment, recruiters will inform grantees of the following:

- The pilot test is not a monitoring review. The focus of the pilot test is on testing the effectiveness of the monitoring tools, methodology, and staff training procedures of the new monitoring system as well as the quality of the collected data; not on the compliance of the grantee.
 - However, the recruiter will also inform the grantee that the review team will be required to report to OHS any issues posing an immediate threat to the health and safety of the children and families served by their program (e.g., supervision, inappropriate release, discipline) which could result in a separate Special Review for the grantee.
- Participation in the pilot test is voluntary. The grantee can refuse to participate in the pilot test with no impact to the status of its current or future grants. Grantees can also withdraw from the pilot test at any time.
- Grantees have the option of participating in one or both pilot test review types (i.e., FA1 and/or FA2).

As part of the incentive plan for maintaining high recruitment rates, recruiters will also inform grantees of the following:

- Pilot testing is an important component of the monitoring system process and that the grantee is part of a special, select group of Head Start programs that have been chosen to contribute to the development of monitoring system.
 - As a token of appreciation for their participation in the pilot test, OHS will provide them with a certificate recognizing contributions to the development of the new monitoring system. Note that the contractor team will prepare the certificate to be issued, but ideally the certificate will come from OHS.
- After the pilot test, project staff will schedule a 30-minute meeting with the grantee to discuss a summary of the information collected during the pilot test to inform the grantee's continuous improvement efforts and to answer any follow-up questions from the grantee.² Note that these meetings will need to be scheduled according to the availability of project staff as well as the convenience of the grantees' schedules.

² Note that the team does not plan on providing grantee review reports to the Pilot Test grantee participants.

The Validation Team will collaborate with the Recruitment Team to track the progress of the recruitment activities, monitor the development of the pilot test sample (including the number of grantees in each project performance-level category that are successfully recruited), and identify and implement any interventions if the pilot test sample becomes skewed. Possible interventions include revising our recruitment messaging to those grantees (e.g., if we identify issues or questions that grantees have) or increasing our incentives.

For grantees that are successfully recruited, the Recruitment Team will schedule the pilot test review in accordance with the grantee's availability and the pilot test review calendar.

Pilot Test Methodology

The following section provides an overview of the methodology of the FA1 and FA2 Pilot Tests. The Pilot Test staff will be comprised of DLH review leads, DLH reviewers, and OHS central office staff. The DLH project team will collaborate with OHS to identify OHS central staff to participate in the Pilot Test.

FA1 Pilot Test

The FA1 protocol guides and EAS will be programmed into IT-AMS; the Pilot Test staff will conduct the test according to the protocol guides and work flow outlined in the FA1 protocol methodology. As in the current FA1 methodology, the FA1 field test will be conducted virtually by one review lead. The Pilot Test staff will conduct the telephone interviews according to the protocol guide instructions, record the responses in IT-AMS, and use those responses to complete the QM ratings. During the FA1 Pilot Test, DLH will hold regular meetings with Pilot Test reviewers, field managers, OHS, and IT staff to update the team on the status of the pilot test, review the preliminary data and address any issues identified in the field or in the data.

Information on the data collection process will be collected from both the respondent and the Feasibility Test staff conducting the Feasibility Test review:

- *Respondent Experience Data:* At the end of each session, the reviewer will ask respondent if there were any questions that the respondent did not understand or needed clarification and why and record their responses.
- *Feasibility Test Staff Experience Data:* The Feasibility Test staff will record information on the work flow, item wording, rating anchors, and review materials (including manual) (e.g., what did not work or was unclear and why) through a questionnaire/notes/diary. Feasibility Test staff will also record timing estimates for implementing the guides and completing the ratings.

Notes and EAS data (e.g., QM rating data) will be collected in IT-AMS and used in analyses to test and finalize the scoring algorithms (including selection between the two draft scoring approaches – The Scoring Calculator (TSC) and the Performance Assessment Matrix (PAM)). These analyses will include:

- Descriptive statistics (e.g., variance estimations)
- Sensitivity analyses and stress testing
 - To test the algorithms ability to distinguish between programs of different levels of performance
- Calibration with OHS review of evidence and resultant scoring
- Potential development of weights

Note that the validation team is exploring including a test of the inter-rater reliability of these measures.

Once the FA1 protocols have been revised and the scoring algorithms have been finalized based on the information collected from the Stage 2 Pilot Test, the EAS development team will meet with OHS to walk through the final protocols, guides, and EAS to describe the features and answer questions. Once OHS approves the review materials, the EAS development team will complete the technical and business requirements for programming into IT-AMS in preparation for the FY2023 monitoring year.

See **Table 2** for an overview of the Stage 2 Pilot Test schedule for FA1.³

Table 2: Stage 2 Pilot Test Schedule for FA1

Task Description	Completion Date(s)
Training of Reviewers on FA1, scheduling, data collection guides (interviews) in IT-AMS and methodology	1/11/22-1/17/22
Conduct pilot test (PT) of FA1 data collection guides (i.e., interviews), EAS (PMs, QIs, QMs), EAS scoring, GPMS scoring	1/24/22-4/1/22
Analyze PT data and revise FY2023 FA1 data collection guides (interviews), EAS (PMs, QIs, QMs), EAS scoring, GPMS scoring based on pilot test data	3/25/22-4/25/22
Submit draft FA1 monitoring protocols, data collection guides (interviews), EAS (PMs, QIs, QMs), EAS scoring, GPMS scoring to OHS for approval, grantee report template	5/02/22
Meetings with OHS to walk through FA1 protocol, data collection guides, EAS (PMs, QIs, QMs) and answer questions	5/03/22-5/16/22
Approval from OHS on FA1 protocol, data collection guides (interviews, etc), EAS (PMs, QIs, QMs)	5/17/22

FA2 Pilot Test

Similar to the FA1 Pilot Test described above, the FA2 protocol guides and EAS will be programmed into IT-AMS and the Pilot Test staff will conduct the test according to the protocol guides and work flow for a typical FA2 review. Accordingly, the FA2 field test will be conducted on-site by a review team, typically comprised of one review lead and three

³ Note that these dates are based on assumptions as of May, 28, 2021. The DLH team will continually review this timeline throughout the development and testing period to report status of project activities, identify and mitigate risk to the schedule, and develop alternatives for OHS to consider if events impact the schedule.

reviewers.⁴⁵ The Pilot Test staff will conduct the data tours, interviews, child file and other record reviews, and explorations according to the protocol guide instructions, record the FA2 monitoring data in IT-AMS, and use those responses to complete the QM ratings. During the FA2 Pilot Test, DLH will hold regular meetings with Pilot Test reviewers, field managers, OHS, and IT staff to update the team on the status of the pilot test, review the preliminary data and address any issues identified in the field or in the data.

Information on the data collection process will be collected from both the respondent and the Feasibility Test staff conducting the Feasibility Test review:

- *Respondent Experience Data:* At the end of each session, the reviewer will ask respondent if there were any questions that the respondent did not understand or needed clarification and why and record their responses.
- *Feasibility Test Staff Experience Data:* The Feasibility Test staff will record information on the work flow, item wording, rating anchors, and review materials (including manual) (e.g., what did not work or was unclear and why) through a questionnaire/notes/diary. Feasibility Test staff will also record timing estimates for implementing the guides and completing the ratings.

Notes and EAS data (e.g., QM rating data) will be collected in IT-AMS and used in analyses to test and finalize the scoring algorithms (including selection between the two draft scoring approaches – The Scoring Calculator (TSC) and the Performance Assessment Matrix (PAM)). These analyses will include:

- Descriptive statistics (e.g., variance estimations)
- Sensitivity analyses and stress testing
 - To test the algorithms ability to distinguish between programs of different levels of performance
- Calibration with OHS review of evidence and resultant scoring
- Development of weights

Note that the validation team is exploring including a test of the inter-rater reliability of these measures.

The collected information will be used by program team (led by Marisa and Melissa) to refine work flow, item wording, rating anchors, and review materials (including the review manuals)

⁴ This is assuming that on-site reviews are permitted by OHS given the most current information available on Covid-19-related travel restrictions. If on-site reviews are not permitted, the DLH monitoring support team will collaborate with OHS to determine the most appropriate alternatives within the parameters of any Covid-19-related travel restrictions.

⁵ Note that the DLH team is exploring whether a portion of the FA2 Pilot Test can be conducted virtually for budget considerations. The team is also exploring including a *mode* test to the Stage 2 Pilot Test design for FA2 to explore whether some FA2 data collection components (e.g., interviews or child file reviews) could be conducted virtually for budget considerations.

prior to finalizing technical/business requirements for IT-AMS programming. See **Table 3** for an overview of the Stage 2 Pilot Test schedule for FA2.⁶

Table 3: Stage 2 Pilot Test Schedule for FA2

Task Description	Completion Date(s)
Training of Reviewers and Review Leads on FA2 data collection guides (RL responsibilities, interviews, data tours, explorations, file and document reviews, etc.) in IT-AMS and methodology	1/24/22-2/18/22
Conduct pilot test (PT) of FA2 data collection guides (interviews, data tours, explorations, file and document reviews, etc.), EAS (PMs, QIs, QMs), EAS scoring, GPMS scoring	2/21/22-4/22/22
Analyze PT data and revise FY2023 FA2 data collection guides (interviews, data tours, explorations, file and document reviews, etc.), EAS (PMs, QIs, QMs), EAS scoring, GPMS scoring based on pilot test data	4/25/22-5/06/22
Submit draft FA2 monitoring protocols, data collection guides (interviews, data tours, explorations, file and document reviews, etc.), EAS (PMs, QIs, QMs), EAS scoring, GPMS scoring, grantee report template to OHS for approval	5/06/22
Meetings with OHS to walk through FA1 protocol, data collection guides, EAS (PMs, QIs, QMs) and answer questions	5/09/22-5/20/22
Approval from OHS on FA1 protocol, data collection guides (interviews, etc), EAS (PMs, QIs, QMs)	5/23/22

⁶ Note that these dates are based on assumptions as of May, 28, 2021. The DLH team will continually review this timeline throughout the development and testing period to report status of project activities, identify and mitigate risk to the schedule, and develop alternatives for OHS to consider if events impact the schedule.

Appendix A: Stage 2 Pilot Test Sampling Approach – Power Analysis

To develop the Stage 2 Pilot Test sampling approach, the validation team conducted power analyses using Cohen’s power tables (Cohen, 1988) for correlational analyses.⁷ Our power analyses were based on the following parameters:

- **Effect size**, in this case, is the estimated magnitude of the correlation coefficient our data analyses will be able to detect with significance. For example, an ES=.20 means that we would be able detect a small relationship with a correlation $r=.20$. **So smaller effect size values mean our sample size allow us to detect smaller significant relationships.** Based on Dr. Fiene’s experience with licensing data, we expect to see larger correlation coefficients (i.e., effect sizes) in our pilot test data – around the magnitude of $r = .40-.50$.
- **Power** estimates the likelihood of our analyses detecting the significant relationship. So .85 indicates we have an 85% chance of detecting a significant relationship. **So larger power values mean we have a greater chance of detecting those relationships.**
- **Alpha** is the level of significance for our statistical tests. For the purposes of these power analyses, Alpha = .05 which is standard.

In addition, a primary consideration of the sampling approach is achieving sufficient variance in the QM ratings, which is critical to the success of this pilot test. That is, we cannot have all grantees with QM ratings=3 in the pilot test sample. As a result, sampling strata will include projected grantee performance levels (low-medium-high), as well as other relevant grantee characteristics (e.g., region, program option).

Table A-1: Power Analysis Chart (based on Cohen 1988)

Option	Sample Size	Effect size (ES)	Power	Alpha (non-directional)
1	52	.40	.85	.05
2	32	.50	.85	.05
3	46	.40	.80	.05
4	28	.50	.80	.05

Table A-1 summarizes the results of our power analysis. Option 3 allows us to specify ~15 grantees in each performance level (i.e., 15 low-performing grantees; 15 medium-performing grantees; 15 high-performing grantees) per the Validation Team’s recommendation. Option 3 also specifies a sample that has an 80 percent likelihood of detecting a correlation

⁷ Cohen, J. (1988). *Statistical power analysis for the behavioral sciences, 2nd ed.* Hillsdale, NJ: Erlbaum. See Chapter 3: The Significance of a Product Moment r_s .

coefficient of $r = .40$ with a significance level of $\alpha = .05$. This results an FA1 pilot test sample size of $N=45$ and FA2 pilot test sample size of $N=45$ for a total of 90 grantees to pilot test both instruments. Note that the Validation Team is exploring the potential for overlap between the FA1 and FA2 pilot test samples to reduce the overall sample.

The Validation Team is also currently exploring using extant information related to expected performance ratings (e.g., findings from past reviews, grantee characteristics, regional office intel) to guide the strata development.

Note that we need to be deliberate about defining the pilot test universe (i.e., which and what type of grantees are eligible to be sampled). The Validation Team recommends being as inclusive as possible (i.e., including Region 11 and 12 grantees) for multiple reasons, including generalizability and performance range. While we recognize that OHS may make a policy decision to exclude certain types of grantees from the GPMS system, that is a separate issue than the pilot test and we would recommend having the pilot test data results inform that policy decision (rather than the current policy drive our pilot test design). Note that OHS has determined that they do not want the pilot test sample to include any grantees that are included in the FY2022 national monitoring schedule.

Lastly, low recruitment rates will have an impact on our sample size parameters (**Table A-2** below). For example, if we only achieve a 50 percent recruitment rate, power shifts to .70 for an effect size of .50. In other words, compared to our target sample size, we would have a lower likelihood of detecting a larger correlation coefficient. Incentives (e.g., feedback for continuous improvement based on information collected through pilot test; gift cards; books and/or toys for the center) may help improve recruitment rates but those obviously are an additional cost.

Table A-2: Recruitment Rate Analysis Chart (based on Cohen 1988)

Recruitment Rate	Sample Size	Effect size (ES)	Power	Alpha (non-directional)
100%	46	.40	.80	.05
~75%	34	.40	.66	.05
~66%	30	.40	.60	.05
~50%	23	.50	.70	.05
~33%	15	.60	.70	.05