# Supporting and Strengthening the Home Visiting Workforce (SAS-HV): Testing and Validation of a Draft Measure of Reflective Supervision for Home Visiting

## Pre-testing of Evaluation Data Collection Activities

## 0970 - 0355

# Supporting Statement

# Part B

**January 2024**

Submitted By:
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

4th Floor, Mary E. Switzer Building
330 C Street, SW
Washington, D.C. 20201

Project Officers:
Nicole Denmark
Shirley Adelstein

# Part B

### B1.    Objectives

*Study Objectives*

The purpose of the Supporting and Strengthening the Home Visiting Workforce (SAS-HV) project is to advance understanding of how to support and strengthen the early childhood home visiting workforce. One focal area of the project is the use of reflective supervision in early childhood home visiting. The Maternal Infant Early Childhood Home Visiting (MIECHV) legislation mandates federally funded home visiting programs maintain high-quality supervision practices, but relatively little is known about specific supervision practices used in the field. This research seeks to address a key gap by developing and testing a measure of reflective supervision that is practice-relevant and useful for research.

Prior phases of this project developed a conceptual model of reflective supervision; reviewed current research, measures, and practice; completed concept mapping data collection and solicited input from the technical workgroup (TWG), practitioner workgroup (PWG), and home visiting model representatives; and developed a draft set of items and response options for the measure.[1] This draft measure was pretested with a sample of home visiting supervisors and revised based on preliminary descriptive results.[2]

The objective of the current phase of the study is to conduct a mixed methods testing and validation study of the draft reflective supervision measure. The data collected will help assess the extent to which the measure is valid and reliable for use in home visiting contexts. At the end of the study, we will produce a technical report (including a manual) summarizing the process and results of testing and validation activities. We will also have a measure that has been revised based on the findings and available for use by researchers and practitioners—however, we will not have tested the revised measure.

*Generalizability of Results*

This study is intended to produce a supervisor self-report measure of reflective supervision that has been tested for (1) factor structure and (2) whether specific claims about the measure (such as its ability to document the presence of reflective supervision techniques in supervision sessions with home visitors) are valid, reliable, and fair. The study is not intended to promote statistical generalization to other service populations. However, the study is intended to provide preliminary evidence that the measure of reflective supervision is appropriate for a broad range of home visiting supervisors with characteristics and contexts consistent with those who participate in testing and validation activities.

*Appropriateness of Study Design and Methods for Planned Uses*

First, the study team will conduct qualitative focus groups with a racially and culturally diverse group of home visiting supervisors to explore the measure's relevance across these subgroups. The semi-

---

[1] Information collection activities were approved by OMB under OMB #0970-0355 on August 19, 2022, with the title Supporting and Strengthening the Home Visiting Workforce (SAS-HV).

[2] Information collection activities were approved by OMB under OMB #0970-0355 on June 12, 2023, with the title Supporting and Strengthening the Home Visiting Workforce (SAS-HV): Online Pretest of Draft Reflective Supervision Measure.

structured focus group guides will include questions related to a) if supervisors feel their racial or ethnic identity influences reflective supervision practices, b) the relative importance of items included in the measure in relation to the racial/ethnic identity of supervisors and dynamics between home visitors and supervisors of varying racial/ethnic identities, b) perspectives on whether the language and terms used in the measure are relevant across subgroups and intergroup dynamics within supervision and c) if items capture racially or ethnically salient practices and techniques in relation to reflective supervision. The use of targeted recruitment strategies and qualitative methods to explore the perceptions of the utility and relevance of the measure for subgroups of home visiting supervisors that identify as Black, Hispanic/Latine, or American Indian and Alaska Native is appropriate for uncovering potentially nuanced and varied perspectives on items within the draft measure. Results from this initial qualitative approach will inform revisions to the measure prior to the developmental test.

The study team will also recruit a developmental sample of approximately 500 home visiting supervisors to complete the measure, using the results to assess item performance, factor structure, internal consistency, validity, and reliability.[3] The use of a sample of this size is needed for analyses to assess the measure's performance (e.g., exploratory and confirmatory factor analyses) and to collect feedback from a broad range of home visiting supervisors on the clarity, acceptability, and perceived value of the measure.

A subsample of supervisors (N= 40) from the developmental test will be included, along with their supervisees (N = 120), in a repeated measures sample. Supervisors in the repeated measures sample will complete a series of web-based quantitative surveys for examining variability in scores on the measure across sessions and supervisees. Supervisees of the repeated measures supervisors' sample will complete a onetime survey asking about the nature and quality of their supervision. This approach is appropriate for yielding a sufficient number of observations for exploring the nature and extent of variability in supervisory practices for a given supervisor and to explore associations between supervisor self-reports and supervisee reports for instrument validation purposes. A subsample of supervisors (N= 15) from the repeated measures sample will also participate in qualitative focus groups to review and interpret the results.

As noted in Supporting Statement A, this information is not intended to be used as the principal basis for public policy decisions and is not expected to meet the threshold of influential or highly influential scientific information.

## B2.    Methods and Design

For all data collection activities included in this request, we propose to gather information from home visiting supervisors implementing home visiting models that are eligible for MIECHV funding (including Tribal MIECHV funds) and are implemented in the United States. To be eligible to take part in either quantitative or qualitative study components, participants must a) supervise home visitors implementing models eligible for MIECHV funds (including Tribal MIECHV funds) and b) must feel comfortable taking part in study activities in English.

The full universe of home visiting supervisors is not known, although estimates suggest more than 2,400 home visiting supervisors[4] are currently working in the field. While a large number of home visiting

---

[3] See the Glossary for definitions of the terminology used in the testing and validation activities.
[4] National Home Visiting Resource Center. (2022). *2022 Home Visiting Yearbook*. James Bell Associates and the Urban Institute.

supervisors will be recruited for study activities, we will be unable to assess how representative our samples are of the total home visiting supervisor population. Instead, we will aim to recruit a sample of home visiting supervisors that is diverse in terms of models implemented, program characteristics, and supervisor and home visitor characteristics.

### B2.1 Sampling Plan and Respondent Recruitment for Focus Groups Exploring the Reflective Supervision Measure's Relevance Across Subgroups

**Sample description:** We will recruit approximately 45 supervisors who identify as Black, Hispanic/Latine, or American Indian or Alaska Native for exploring the relevance of the measure across racial and ethnic subgroups. This sample size will allow us to conduct up to three focus groups for each of these three subgroups of participants. Each focus group will be comprised of approximately five participants, for a total of 15 participants within each subgroup. Our sample size estimations are based on guidance regarding best practices in conducting focus groups in person and virtually (Daniels et al., 2019; Guest et al., 2017; Hennink et al., 2019).

**Recruitment plan:** We plan to initially recruit participants for these focus groups by engaging existing workgroups of diverse home visiting supervisors to identify interested home visiting supervisors. Existing workgroups include a group of racially and ethnically diverse supervisors that are currently exploring the cultural relevance of reflective supervision and a workgroup comprised of racially and ethnically diverse supervisors and home visitors. We will also recruit participants through the Home Visiting Applied Research Collaborative's (HARC) Practice-Based Research Network (PBRN)[5] newsletter and announcements in other relevant newsletters (e.g., the National Home Visiting Resource Center, Office of Planning, Research and Evaluation (OPRE), Health Resources and Services Administration (HRSA), Tribal Home Visiting). See Appendix A for the recruitment announcement and Appendix B for the participant confirmation and preparation email.

Finally, if desired sample sizes are not obtained through these methods, we will work with recruited participants to engage in snowball recruitment of additional participants. Because participants will be purposively selected, they will not be representative of the population of supervisors within the home visiting field.

### B2.2 Sampling Plan and Respondent Recruitment for Developmental Sample

Quantitative sample

**Sample description:** We will aim to recruit about 500 supervisors for our developmental sample. Our proposed sample size follows recommended guidelines, with a sample of 300 considered typically adequate for exploratory factor analysis (EFA), and a sample of approximately 200 considered adequate for confirmatory factor analysis (CFA) (DeVellis & Thorpe, 2022). EFA and CFA provide stable solutions most reliably with sufficiently large samples (Hogarty et al., 2005), and exploratory and confirmatory analyses must be conducted on independent samples (i.e., randomly splitting the developmental sample to conduct first EFA, followed by CFA). Sample size estimation for factor analysis is complex because it depends on characteristics of the data, such as the number of factors and items per factor, magnitude of communalities, magnitude of factor loadings, and data missingness (Harrington, 2009; McNeish, 2017).

---

[5] The purpose of HARC's PBRN is to provide a national network for conducting collaborative, field-initiated studies with local home visiting programs, regardless of the model of home visiting being used. The PBRN is a voluntary network of hundreds of local home visiting programs, including individual sites and state and local networks.

Larger samples are required for EFA because assumptions about the factor structure and the size of loadings cannot be made in advance.

**Recruitment plan:** As a centralized, national list of home visiting supervisors does not exist, recruitment of home visiting supervisors will occur through MIECHV leads and model representatives.[6]
We aim to obtain a sample that is diverse in terms of supervisor race and ethnicity. We will also aim to obtain diversity in other supervisor characteristics (e.g., length of time providing supervision, training and professional development in reflective supervision), characteristics of supervisees and families served (e.g., race, ethnicity, language), and program characteristics (e.g., home visiting models implemented, model requirements for reflective supervision, type of agency home visiting program is situated within, size of home visiting program, region of the country).

Recruitment activities will include:

- Working with the Contracting Officer's Representative (COR)[7] and HRSA federal project officers to compile a list of all state, territory, and Tribal MIECHV leads (hereafter referred to MIECHV leads). Project officers will be copied on all outreach and communication with MIECHV leads if that is preferred by the federal project officer.

- Working with the National Alliance of Home Visiting Models[8] to obtain contact information for home visiting models implemented in the United States and eligible for MIECHV funds. We will also engage in direct outreach to models that are not represented in the alliance.

- Hosting an informational webinar for MIECHV leads and model representatives to introduce the study and upcoming outreach and recruitment efforts. The webinar will be optional to attend and recorded for later viewing.

- Conducting direct email outreach to all MIECHV leads and model representatives, including sharing information about the study and requesting assistance in distributing recruitment materials through their recommended channels (i.e. sharing an existing list of email addresses, adding to regular communication products like newsletters). We will engage in targeted outreach to Tribal MIECHV grantees and states with Tribal local implementing agencies (LIAs) to ensure AIAN supervisors receive notification of the study and are encouraged to participate. All outreach materials will be electronic PDFs and links to online information that can be forwarded by MIECHV leads and models to primary points of contact at local programs. Outreach and recruitment materials may be adapted and tailored to reach diverse groups of supervisors and encourage study participation among diverse groups of supervisors.

---

[6] Each MIECHV state, territory, and tribal awardee has a primary point person, or lead. MIECHV awardees select one or more models from a HRSA-approved list of evidence-based home visiting models. Model implementation is supported by model-specific organizations that provide training and technical assistance and fidelity criteria. Representatives from these home visiting model-specific organizations will be one avenue for home visiting supervisor outreach.

[7] A contracting officer's representative (COR), a federal government staff person, assists in the administration or technical monitoring or administration of a contract.

[8] The National Alliance of Home Visiting Models is a collaboration of nine home visiting program models. See list of participating models here: https://www.nationalalliancehvmodels.org/models

- Offering the option of one-on-one phone calls with MIECHV leads and model representatives to address concerns and answer questions. We will also ask permission to reach back out to MIECHV leads and models to assist in targeted recruitment of supervisors with certain characteristics based on response rates.
- If preferred by MIECHV leads and models, sharing outreach and recruitment materials directly with local programs.
- Monitoring response rates and participant characteristics to inform ongoing targeted outreach and recruitment as necessary. For example, if a particular model or participant characteristic appears to be underrepresented, we will reach out to MIECHV leads and models to ask for their assistance in identifying supervisors implementing those models or with underrepresented characteristics.

The web-based survey with the developmental sample will include a question asking if the participant gives permission for the study team to reach back out to them for subsequent data collection activities, including completion of the reflective supervision measure at multiple time points and follow up focus groups. See Appendix C for recruitment materials.

Repeated measures supervisors subsample and supervisees sample

**Sample description:** We will recruit a sample of approximately 40 supervisors from the developmental sample to (1) engage their supervisees in the study (N =120) and (2) complete the reflective supervision measure at multiple time points after providing reflective supervision to the same home visitor (e.g., up to 3 times). We will ask each supervisor to repeat this process for up to 3 home visitors, resulting in an estimated 360 total observations over time. Each supervisee will be asked to complete a one-time survey on the nature, quality, and satisfaction with their supervision. These sample sizes and number of observations are adequate for exploring associations between supervisor and supervisee reports and exploring the nature and extent of variability in supervisory practices for a given supervisor across supervision sessions and across supervisees.

To participate in a repeated measures data collection, supervisors must a) provide at least bi-weekly supervision to at least three home visitors b) indicate that they are willing to complete all study activities over a period of approximately 8 weeks and c) seek voluntary participation by up to three home visitors they supervise. A primary goal of this work is to examine within-person variability over time and across home visitors; thus, given time constraints and potential participant burden, we will use convenience sampling to identify supervisors who are willing to participate in this activity, making attempts to recruit a sample diverse in supervisor, home visitor, and program characteristics.

**Recruitment plan:** We will select prospective supervisor participants based on results of preliminary quantitative analyses and whether participants have indicated willingness to take part in this activity on the developmental testing survey. We will offer a informational webinars on a rolling basis for selected supervisors and their supervisees to review the procedures for completing the measure at multiple time points and steps for supervisees to complete a survey. Supervisors participating in the repeated measures will be assigned a study liaison who will support their participation and submission of the reflective supervision measure at multiple time points. This can include reminders for sessions and survey completion, and answering questions as they arise. See Appendices D and E for recruitment materials.

Qualitative subsample

**Sample description:** We will identify a subsample of approximately 15 supervisors who were part of the repeated measures sample. This sample size will allow us to conduct up to three focus groups, each comprised of 5 participants. Our sample size estimations are based on guidance regarding best practices in conducting focus groups in person and virtually (Daniels et al., 2019; Guest et al., 2017; Hennink et al., 2019). To be eligible to take part in this activity, participants must have a) completed the quantitative survey and repeated measures described above, and b) indicated willingness to take part in a focus group.

**Recruitment plan:** We will select participants based on the results of preliminary quantitative analyses. If a participant agrees and they are selected for participation in a focus group, we will reach out to them directly via email. See Appendix F for recruitment materials and Appendix G for the participant confirmation and preparation email.

**B3.     Design of Data Collection Instruments**

*Development of Qualitative Data Collection Instruments*

The focus group protocol (Instrument 2) for exploring relevance across subgroups will use a semi-structured qualitative approach (for example, Morgan, 1996). The protocol includes a set of guiding questions, optional probes, and allowance for the facilitator to further probe and tailor questions based on participants' responses. The protocol is divided into sections to understand participants' overall reflective supervision practices in the context of their racial or ethnic identity, as well as a section to understand participants' feedback on a subsample of items from the reflective supervision measure. This approach is designed to gather information on culturally relevant reflective supervision practices and techniques, as well as on how items in the measure are salient and applicable across different subgroups.

The focus group protocol (Instrument 8) will involve reviewing and interpreting survey results with a subsample of the supervisors completing the repeated measures. The protocol includes example questions that we will use to guide an open discussion. We will further refine questions based on the quantitative survey results. We anticipate the questions will cover the supervisors' experiences and perspectives with completing the measure multiple times (i.e. the measure's accuracy in capturing the nature of sessions, its length, and feasibility for use in practice), discussion of concordance between supervisor and supervisee results, and implications for next steps.

*Development of Data Collection Instruments for Quantitative Testing*

Data collection for developmental testing includes completing four instruments within one online survey: a) the participant and contextual characteristics questionnaire, b) the reflective supervision measure, c) participant perspectives on the measure, and d) measures to examine convergent and concurrent validity.

The development of the draft reflective supervision measure (Instrument 3) was guided by a systematic review of existing elements of reflective supervision and informed by the Study Team's knowledge of existing research and literature on reflective supervision. Elements of reflective supervision represented in the measure were first refined in collaboration with technical and practitioner workgroup members. The Study Team then conducted a concept mapping process with practitioners and researchers, including a group interpretation meeting. Informal cognitive interviews were conducted to identify

issues with interpretation, definitions, and response options, and a small web-based pretest was conducted to guide further revisions.

The participant characteristics questionnaire (Instrument 1) and participant perspectives questionnaire (Instrument 4) were developed for the pretesting process in the most recent phase of the work and will be replicated for continuity in this phase.

Measures to examine validity (Instrument 5) were selected from a literature review and by soliciting recommendations from a subset of TWG members. We selected established supervisor self-report measures of constructs that we could expect to correlate with the draft reflective supervision measure when administered at the same time, including supervisory style, reflective supervision self-efficacy, and supervisory working alliance. We selected specific measures based on the extent to which they were theoretically sound and demonstrated some evidence of content and face validity, internal consistency reliability, and criterion validity with related measures when tested in similar contexts (e.g., health and human services). Due to the paucity of available measures that met these criteria, we also added a small set of single-item, exploratory questions drawn and/or adapted from related literature to assess goals for supervision, commitment to a reflective supervision approach, and self-rating of reflective supervision skills.

*Development of Data Collection Instruments for Repeated Measures*

Lastly, the repeated measures supervisor subsample will complete a brief questionnaire (Instrument 6) about the nature of the supervision sessions they are reporting on and the home visitors they supervise. This instrument was developed by the Study Team to collect needed contextual information on each home visiting session where the measure of reflective supervision was completed.

Supervisee measures (Instrument 7) were selected from a literature review and by soliciting recommendations from a subset of TWG members. We selected established supervisee self-report measures of constructs that we could expect to correlate with the draft reflective supervision measure, including supervisory style, reflective supervision quality, supervisory relationship, supervisory satisfaction, and self-reflection and insight. We selected specific measures based on the extent to which they were theoretically sound and demonstrated some evidence of content and face validity, internal consistency reliability, and criterion validity with related measures when tested in similar contexts (e.g., health and human services).

## B4.     Collection of Data and Quality Control

The Study Team will collect data from home visiting supervisors and supervisees that consent to participate in each information collection activity. Recruitment of supervisors into the study will involve outreach to potential participants described in detail above in section B2. Recruitment materials are included in the Appendices.

The developmental testing data collection instruments will be programmed into one survey in Qualtrics, a web-based survey software. To monitor the survey data for quality and consistency, the Study Team will review reports on survey completion rates and item completion rates periodically during the data collection period.

ACF has contracted with James Bell Associates and partners Johns Hopkin University and University of Colorado Anschutz Medical Campus to conduct the study. James Bell Associates will lead all data

collection and analysis activities in collaboration with Johns Hopkin University and University of Colorado Anschutz Medical Campus.

## B5.    Response Rates and Potential Nonresponse Bias

*Response Rates*

The study team used research-based guidelines to determine the developmental testing sample sizes recommended for conducting the intended analyses (see sample descriptions in section B2 above).  It may be challenging to recruit the sample size (of approximately 500 home visiting supervisors) within the two-month timeframe for study recruitment. To minimize this challenge, we are planning multiple recruitment avenues and steps. We will also explore recruiting supervisors from promising models (models not yet meeting the evidence standards required for MIECHV funding and that may not be represented in the model alliance described above) to obtain necessary sample sizes.

We aim to obtain a sample that is diverse in terms of supervisor race and ethnicity. We will engage in targeted outreach to increase these percentages as much as possible. We will also aim to obtain diversity in other supervisor characteristics (e.g., length of time providing supervision, training and professional development in reflective supervision), characteristics of supervisees and families served (e.g., race, ethnicity, language), and program characteristics (e.g., home visiting models implemented, model requirements for reflective supervision, type of agency home visiting program is situated within, size of home visiting program, region of the country). We will monitor response rates and participant characteristics to inform targeted outreach and recruitment as necessary.

*Non-response*

As participants will not be randomly sampled and findings are not intended to be representative, nonresponse bias for the web-based measure developmental test will not be calculated. Participant demographics will be documented and reported in written materials associated with the data collection, however, we will not have demographic information on the universe of home visiting supervisors with which to compare our sample.

## B6.   Production of Estimates and Projections

Data will not be used to generate population estimates, either for internal use or dissemination.

## B7.   Data Handling and Analysis

### B7.1 Data Handling

To minimize errors in survey data processing and analysis, we will review all data during initial processing before data is approved and becomes part of the final data set.  Additionally, the Qualtrics web-based software employs quality controls to ensure only valid responses are allowed and to minimize missing responses.

### B7.2 Qualitative Analysis of Focus Groups

We will use a summary-based approach to analyze qualitative data that balances rigor with efficiency (Morgan, 2019). This approach is suitable for applied projects in which results are needed quickly to inform future waves of data collection.

Data reduction, analysis, and interpretation will be iterative, using both predetermined and emergent codes (Hsieh and Shannon 2005).  In the initial round of coding, we will apply predetermined codes that

align with the focus group questions. In the second stage of coding, we will develop additional emergent codes based on discussions within the internal team about emerging patterns and themes. Data interpretation will include going back and forth between transcripts and notes, focus group protocols, and the draft measure of reflective supervision. We will develop tables that organize findings by focus group and by question.

We will look for patterns or themes across focus groups. We will summarize themes and share findings with the COR, the TWG and the PWG to inform possible revisions to the measure and cultural considerations for administration of the measure, to be included in a technical manual.

### B7.3 Quantitative Data Analysis

Analysis of Missing Data

Before beginning other analyses, we will first conduct analysis of missing data of all quantitative data to identify any non-random patterns of missingness (e.g., items that are frequently skipped, missingness related to participant characteristics). These analyses will provide information about the plausibility of assumptions (i.e., missing-at-random) required for planned inferential analytic techniques (i.e., full information maximum likelihood estimation) that can increase power and precision in analyses. We will confirm that underlying assumptions can be reasonably made before proceeding with analyses.

Descriptive Analyses

Our first steps with the quantitative data will be descriptive analysis. We will examine demographic characteristics of participating supervisors (e.g., gender, race/ethnic/cultural background, age) along with descriptors of their experiences with home visiting (e.g., years in practice, home visiting models implemented).

Descriptive analysis for all quantitative measures will include examination of central tendency (mean, mode, median) and dispersion (minimum, maximum, range, standard deviation, univariate normality, distributional skew, floor, and ceiling effects). These steps are necessary to check assumptions underlying inferential analyses and to prepare data for reliability and validity analyses of the measure of reflective supervision. Careful examination of items included in the reflective supervision measure will help identify items that may be 1) ambiguous, 2) incorrectly keyed or scored, 3) extremely skewed, and 4) not discriminative enough, with very low variability (Price, 2017). It will also screen for potential outliers that may need to be excluded from analyses.

If sufficient sample sizes are obtained within subgroups, we will also explore variability related to subgroups based on individual (e.g., ethnicity) and contextual (e.g., home visiting model) characteristics. We will conduct analyses to evaluate subgroup differences using appropriate statistical tests depending on level of measurement, number of categories, and distribution of the data. This will give us insights into whether items on the measure operate differently across subgroups (e.g., load on different factors) and whether associations of factors with variables included for validity analyses differ across groups.

We will examine bivariate correlation matrices of all items within the measure and of these items with other quantitative measures included for validity purposes. This initial scan of associations will be important to understanding the overall covariance matrix that will be the foundation of factor, reliability, and validity analyses described below and will allow us to identify any surprising patterns in associations.

Exploratory Factor Analysis

We will use EFA to examine the structure of the draft measure and identify potential subscales for further examination. EFA is a data-driven approach used to identify latent (unobserved) factors that explain covariation within a set of variables. EFA casts a broad net to evaluate potential relationships among draft items, estimating loadings and cross loadings of every item on every identified factor (all error covariances are constrained at zero in EFA models to allow the model to be estimated).

As described above, we will first ensure that statistical assumptions are plausible (e.g., linearity, multivariate normality) and that the data are appropriate for EFA (e.g., factorability using Bartlett and Kaiser-Meyer-Olkin (KMO) statistics). Our initial examination of the data will inform the selection of appropriate estimation and factor rotation methods.

Determining the best fitting EFA model will be an iterative process. For example, it is possible that the first round of analyses may point to a particular factor solution, but examination of factor loadings may reveal the lack of a simple factor structure (i.e., some items may double-load or strongly cross load on multiple factors, or some "orphan" items may not load on any factors at all). In such cases, we will consider dropping items and running a new analysis with the restricted item pool (decisions to drop items will also be informed by review of the descriptive statistics described above).

Once we have determined a final, optimized factor structure, we will summarize the solution, reporting eigenvalues, factor loadings (pattern coefficient matrix), factor intercorrelations, and relevant fit indices. We will describe how we determined the number of factors to retain and our strategy for interpreting factors.

Confirmatory Factor Analysis

CFA will provide information on the stability and replicability of the model. CFA tests whether a specified model fits the data (i.e., whether or not the covariance matrix estimated based on the model adequately approximates the covariance matrix from the observed data; Harrington, 2009). We will use CFA to test the model identified through the EFA in an independent sample of 200 participants from the developmental sample, namely those not randomized into the EFA sample.

With the final EFA model as our guide, we will specify the factor model to be tested using CFA. We will check model fit using parameters described for testing EFA model fit (i.e., CFI, SRMR, RMSEA, chi-square). If the model demonstrates poor overall fit to the data, we will examine modification indices to identify areas of poor fit to determine whether further refinements to the model should be considered. For example, we may consider allowing for correlated errors to account for method effects (e.g., similar item wording or structure).

If we can recruit a sample of supervisors with sufficient diversity (see recruitment plan in Section B2.2 above), we plan to utilize multigroup CFA to test for invariance across subgroups of supervisors.

Reliability Analysis

We will also calculate and report internal consistency reliability estimates (i.e., Cronbach's alpha) for each identified factor and will examine inter-item and item-scale correlations along with alpha-if-item-deleted statistics to identify items that might be able to be trimmed without effecting the overall reliability of each subscale. Cronbach's alpha values of .70 or higher generally suggest adequate internal consistency, whereas values of .95 or above suggest that some items may be redundant and might be removed to reduce the length of the measure without compromising the overall reliability of the dimension (Tavakol & Dennick, 2011).

Validity Analysis

**Concurrent Validity Analysis:** Once we have confirmed the factor structure of the measure through the analyses described above, we will utilize the full sample (N=500) to assess concurrent validity by examining associations between scores (overall and subscale/dimension) and scores on other measures of closely related constructs administered at the same time. These include existing supervisor-report measures of supervisory relationship quality or clinical or counseling supervision practices (as described in Design of Data Collection Instruments, Section B3, above). The choice of statistical tests for concurrent validity will depend on the format of each measure; we anticipate using Pearson's correlations for these tests when appropriate data assumptions are met (e.g., linearity, normal distribution). Significant correlations between the measure or subscale and related measures will be indicative of concurrent validity.

**Convergent Validity Analysis:** Similarly, we will examine convergent validity by examining associations with validated measures of constructs theoretically related to provision of reflective supervision, such as perceived supervision quality, reflective supervision self-efficacy, reflective capacity and/or mindfulness, and satisfaction with the experience of providing reflective supervision. As with concurrent validity analyses, we anticipate using Pearson's correlations to examine validity here, but will make the final choice of appropriate statistical tests based on examination of the data vis-à-vis underlying statistical assumptions. Significant correlations between the measure or subscale and related measures will be indicative of convergent validity.

Repeated Measures Analysis

Using repeated measures data, we will examine the nature and extent of variability in supervisory practices, as measured by the reflective supervision measure, for a given supervisor across supervision sessions and across supervisees. We will calculate measures of central tendency (mean, mode, median) and dispersion (minimum, maximum, range, standard deviation) for both the overall reflective supervision measure and for any subscales identified through the factor analysis process and will also look closely at individual items to identify any with particularly high or low levels of variability across sessions. We will estimate correlations across time and across home visitors to explore stability on these different dimensions. We will inspect data visually using graphs and scatterplots to explore patterns of variability.

To explore the possibility that reports on the measure may change systematically over time (i.e., due to repeated use of the measure and/or changes in practices with experience), we will examine time-related trends in scores. Because of the small sample size for this component of the study, these analyses will most likely involve repeated measures analysis of variance, although we will explore the feasibility of using linear mixed modelling or longitudinal growth curve modeling with Bayesian (small sample) estimators as well. Analysis of repeated measures data will be conducted using Mplus software version 8.9 (or subsequent version; (Muthén & Muthén, 2017).

To examine associations between supervisor reports using the reflective supervision measure and supervisee reports, we will conduct a series of regression analyses predicting scores on supervisee measures from scores on the repeated RS measures the supervisor has completed for each supervisor. This analysis will be conducted using the "Type=Complex" functionality in Mplus (version 8.9 or subsequent version; Muthén & Muthén, 2017) to account for the clustered nature of the data (by supervisor). A sample size of approximately 40 supervisor-supervisee pairs is appropriate for this type of clustered regression analysis.

*B7.4 Data Use*

Activities completed in this data collection will produce a supervisor self-report measure of reflective supervision that has been tested for (1) factor structure and (2) whether specific claims about the measure (such as its ability to document the presence of reflective supervision techniques in supervision sessions with home visitors) are valid, reliable, and fair. These claims will only be applicable to the sample included in the testing and validation activities. We will also have information regarding contemporaneous associations between scores on the measure and constructs hypothesized to be closely related to reflective supervision and how well the measure captures variability in reflective supervision practices across sessions and across supervisees. At the end of the study period, we will produce a technical report (including a manual) summarizing the process and results of testing and validation activities. We will also have a measure that has been revised based on the findings—however, this project does not include time for testing of the revised the measure.

## B8.  Contact Persons

The information for this study is being collected by James Bell Associates, Johns Hopkin University, and University of Colorado Anschutz Medical Campus on behalf of ACF. Principal Investigator Allison West (awest25@jhu.edu), Project Director Mariel Sparr (sparr@jbassoc.com) and Task Lead Nancy Whitesell (nancy.whitesell@ucdenver.edu) led development of the study design plan and data collection protocols and will oversee collection and analysis of data.

The agency responsible for receiving and approving contract deliverables is:

> The Office of Planning, Research, and Evaluation (OPRE),
> Administration for Children and Families (ACF)
> U.S. Department of Health and Human Services

The Federal project officers for this project are Nicole Denmark and Shirley Adelstein.

**Alternative Supporting Statement for Information Collections Designed for
Research, Public Health Surveillance, and Program Evaluation Purposes**

**Appendices: Recruitment Materials**
**Appendix A**: Recruitment Announcement for Focus Groups to Explore Relevance of Reflective Supervision Measure Across Subgroups
**Appendix B**: Confirmation and Preparation Participant Email for Focus Groups to Explore Relevance of Reflective Supervision Measure Across Subgroups
**Appendix C**: Recruitment Email for Web-based Developmental Test of Reflective Supervision Measure
**Appendix D**: Recruitment Email for Supervisors for the Repeated Measures Subsample
**Appendix E:** Recruitment Email for Home Visitor Survey (for Supervisees of Repeated Measures Supervisors)
**Appendix F**: Recruitment Email for Focus Groups with Repeated Measures Supervisors
**Appendix G**: Confirmation and Preparation Email for Focus Groups with Repeated Measures Supervisors
**Appendix H**: Study FAQs


**Attachments: Study Instruments**
**Instrument 1:** Participant and Contextual Characteristics Questionnaire
**Instrument 2:** Focus Group Protocol Exploring Relevance Among Racial and Ethnic Subgroups
**Instrument 3:** Reflective Supervision Measure
**Instrument 4:** Participant Perspectives of the Reflective Supervision Measure
**Instrument 5**: Measures to Examine Convergent and Concurrent Validity
**Instrument 6:** Supervisor Survey for Repeated Administration of the Reflective Supervision Measure
**Instrument 7:** Home Visitor Survey (for Supervisees of Repeated Measures Supervisors)
**Instrument 8**: Focus Group Protocol for Repeated Measures Supervisor Subsample

**Glossary**

**Factor analysis** is a statistical technique used to identify the optimal number of constructs, called factors, that fit a list of items. Factors are higher level, more abstract concepts than individual items. In **exploratory factor analysis,** variation in item scores is used to assign items to factors to maximize the amount of shared variation and internal consistency for each factor (measured using factor loadings). **Confirmatory factor analysis** is used to test this factor structure with an independent sample. **Eigenvalues** are calculations of the amount of information captured by a factor and are used to assess how well the factor structure condenses the items.

Scale developers use reliability and validity tests to assess the performance of a measure and its factors. Reliability reflects a measure's consistency and replicability, and validity tests the accuracy of a measure (measuring what is intended to be measured). **Internal consistency reliability** (usually estimated with a test called Cronbach's alpha) assesses the extent to which the items within a factor co-vary, with strong correlations the goal. **Test-retest reliability** assesses the consistency of a participant's scores over time. Some variation over time is expected for the reflective supervision measure. **Convergent validity** assesses the degree to which scores (at the factor or measure level) correlate with other measures theoretically related to reflective supervision constructs. Similarly, **concurrent validity** assesses associations between *established* measures related to reflective supervision.

**Alternative Supporting Statement for Information Collections Designed for Research, Public Health Surveillance, and Program Evaluation Purposes**

## References

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in Public Health*, *6*, 149.

Daniels, N., Gillen, P., Casson, K., & Wilson, I. (2019). STEER: Factors to consider when designing online focus groups using audiovisual technology in health research. *International Journal of Qualitative Methods*, *18*, 1609406919885786.

DeVellis, R.F. (2017). *Scale development: Theory and applications*. 4th Edition. Los Angeles: Sage.

DeVellis, R. F. & Thorpe, C.T. (2022). *Scale development: Theory and applications.* 5th Edition. Los Angeles: Sage.

Duggan, Anne, Ximena A. Portilla, Jill H. Filene, Sarah Shea Crowne, Carolyn J. Hill, Helen Lee, and Virginia Knox (2018). *Implementation of Evidence-Based Early Childhood Home Visiting: Results from the Mother and Infant Home Visiting Program Evaluation*, OPRE Report # 2018-76A, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Guest, G., Namey, E., & McKenna, K. (2017). How many focus groups are enough? Building an evidence base for nonprobability sample sizes. *Field Methods*, *29*(1), 3-22.

Efstation, J., Patton, M.J., & Kardash, C. M. (1990). Measuring the working alliance in counselor supervision. *Journal of Counseling Psychology*, *37*, 322-329.

Harrington, D. (2009). *Confirmatory factor analysis.* Oxford University Press.

Hennink, M. M., Kaiser, B. N., & Weber, M. B. (2019). What influences saturation? Estimating sample sizes in focus group research. *Qualitative Health Research*, *29*(10), 1483-1496.

Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement*, *65*(2), 202–226. https://doi.org/10.1177/0013164404267287

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277-1288.

McNeish, D. (2017). Exploratory factor analysis with small samples and missing data. Journal of personality assessment, *99*(6), 637-652.

Morgan, D. L. (1996). Focus groups. *Annual review of sociology*, *22*(1), 129-152.

Morgan, D.L. (2019). *Basic and advanced focus groups.* Sage.

Moyers, T. B., Rowell, L. N., Manuel, J. K., Ernst, D., & Houck, J. M. (2016). The motivational interviewing treatment integrity code (MITI 4): rationale, preliminary reliability, and validity. *Journal of substance abuse treatment*, *65*, 36-42.

Muthén, L. K., & Muthén, B. O. (2017). Mplus: Statistical Analysis with Latent Variables: User's Guide (Version 8). Los Angeles, CA: Authors.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8

National Home Visiting Resource Center. (2022). *2022 Home Visiting Yearbook.* James Bell Associates and the Urban Institute.

Price, L. R. (2017). *Psychometric Methods: Theory into Practice.* New York: The Guilford Press.

Priddis, L. & Rogers, S.L. (2018) Development of the reflective practice questionnaire: preliminary findings. *Reflective Practice*, *19*(1), 89–104.

Roggman, L. A., Cook, G. A., Innocenti, M. S., Jump Norman, V. K., Boyce, L. K., Olson, T. L., ... & Peterson, C. A. (2019). The home visit rating scales: Revised, restructured, and revalidated. *Infant Mental Health Journal*, *40*(3), 315-330.

Rønnestad, H., & Lundquist, K. (2009). *Brief supervisory alliance questionnaire:* Oslo, Norway: Department of Psychology, University of Oslo.

Sandstrom, Heather, Sarah Benatar, Rebecca Peters, Devon Genua, Amelia Coffey, Cary Lou, Shirley Adelstein, and Erica Greenberg. 2020. Home Visiting Career Trajectories: Final Report. OPRE Report #2020- 11, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Shea, S., Goldberg, S. & Weatherston, D. (2012) *Reflective Supervision Self-Efficacy Scale.* Unpublished manuscript.

Szymanski, D. M. (2003). The feminist supervision scale: A rational/theoretical approach. *Psychology of Women Quarterly*, *27*(3), 221-232.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, *2*, 53.

Weiss, R. S. (1995). *Learning from strangers: The art and method of qualitative interview studies*. Simon and Schuster.