

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/317858773>

Flexicution in Program Evaluation: Developing a Two-Tier Plan to Address Diverse Priorities Across the Four Services' Intensive Home Visitation Programming

Article in *Military Behavioral Health* · June 2017

DOI: 10.1080/21635781.2017.1343697

CITATIONS

0

READS

23

5 authors, including:



Tara Saathoff-Wells

Pennsylvania State University

23 PUBLICATIONS 90 CITATIONS

SEE PROFILE



Jennifer Karre

Pennsylvania State University

17 PUBLICATIONS 163 CITATIONS

SEE PROFILE



Katie Davenport

Pennsylvania State University

15 PUBLICATIONS 44 CITATIONS

SEE PROFILE



Daniel Francis Perkins

Pennsylvania State University

209 PUBLICATIONS 3,774 CITATIONS

SEE PROFILE

Rapid #: -19384853

CROSS REF ID: **2797446**

LENDER: **CPE :: University Libraries**

BORROWER: **UPM :: Paterno Library**

TYPE: Article CC:CCL

JOURNAL TITLE: Military behavioral health

USER JOURNAL TITLE: Military Behavioral Health

ARTICLE TITLE: Flexicution in Program Evaluation: Developing a Two-Tier Plan to Address Diverse Priorities Across the Four Services' Intensive Home Visitation Programming

ARTICLE AUTHOR: Tara Saathoff-Wells, Jennifer K. Karre, Katie Dave

VOLUME: 5

ISSUE: 4

MONTH: 6

YEAR: 2017

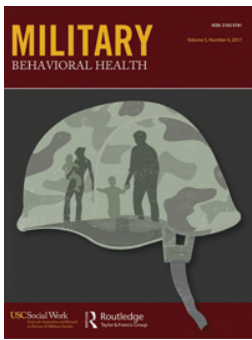
PAGES: 324-334

ISSN: 2163-5781

OCLC #:

Processed by RapidX: 8/10/2022 10:25:57 AM

This material may be protected by copyright law (Title 17 U.S. Code)



Flexicution in Program Evaluation: Developing a Two-Tier Plan to Address Diverse Priorities Across the Four Services' Intensive Home Visitation Programming

Tara Saathoff-Wells, Jennifer K. Karre, Katie Davenport, Mary E. Campise & Daniel F. Perkins

To cite this article: Tara Saathoff-Wells, Jennifer K. Karre, Katie Davenport, Mary E. Campise & Daniel F. Perkins (2017) Flexicution in Program Evaluation: Developing a Two-Tier Plan to Address Diverse Priorities Across the Four Services' Intensive Home Visitation Programming, *Military Behavioral Health*, 5:4, 324-334, DOI: [10.1080/21635781.2017.1343697](https://doi.org/10.1080/21635781.2017.1343697)

To link to this article: <https://doi.org/10.1080/21635781.2017.1343697>



Published online: 27 Jul 2017.



Submit your article to this journal [↗](#)



Article views: 160



View related articles [↗](#)



View Crossmark data [↗](#)

SPECIAL SECTION



Flexicution in Program Evaluation: Developing a Two-Tier Plan to Address Diverse Priorities Across the Four Services' Intensive Home Visitation Programming

Tara Saathoff-Wells^a, Jennifer K. Karre^a, Katie Davenport^a, Mary E. Campise^b, and Daniel F. Perkins^{a,c}

^aClearinghouse for Military Family Readiness at Penn State, The Pennsylvania State University, University Park, Pennsylvania; ^bMilitary Community and Family Policy, Office of the Secretary of Defense, Alexandria, Virginia; ^cDepartment of Agricultural Economics, Sociology, and Education, The Pennsylvania State University, University Park, Pennsylvania

ABSTRACT

The New Parent Support Program (NPSP) is an intensive secondary prevention, parent-education program that serves high-needs families with very young children in the Active Components of the Army, Air Force, Navy, and Marine Corps. The Clearinghouse for Military Family Readiness at Penn State (Clearinghouse) was tasked to work with all of the armed services to develop a common logic model and develop and test an expanded evaluation plan for NPSP. Using the principles of action research and practical program evaluation, the collaborative approach ensured that each service branch participated in building a logic model and evaluation plan that accurately represented their respective strengths, needs, and mission priorities. Program managers at the command levels of each service and the program manager at the Department of Defense level were key partners in working with Clearinghouse evaluators. The result is a two-tier model that will be implemented in a multisite evaluation setting. This model builds on common practices and measures and will offer flexibility for future growth, unique service priorities, and evaluation capacity.

KEYWORDS

Service providers; parenthood/parent–adult child relations; families and children; resilience; evidence-based practice

The New Parent Support Program (NPSP) is a secondary prevention, parent education program that is well-established across the Active Component of all four United States Military Services (i.e., Army, Air Force, Navy, Marine Corps). NPSP is delivered through home visitation services, and its intent is to prevent child maltreatment in high-needs families (DoDI 6500.05; Department of Defense [DoD], 2012). This program presented an opportunity to create and implement a multisite evaluation (MSE) to better understand how, when, under what conditions, and for which participants NPSP works. Multisite evaluations present unique and often complex challenges that must be addressed early, monitored, and managed throughout the evaluation process. Real-world conditions impact consistency in delivery, emphasize variations between sites' outcome priorities, uncover differences in target populations served, and highlight resource allocation variations for program operations and sustainability. In short, all sites are rarely equal in implementation.

This article first describes NPSP and the scope of the evaluation project. The second section discusses action research and participatory program evaluation as the framework for building and conducting an MSE for

NPSP. The third section identifies the key elements of designing and implementing an MSE that could accommodate service branch differences yet yield a common data set and discusses how these elements are manifested in this project. The final portion of this article demonstrates how these elements were important in guiding the two-tier program evaluation plan and its pilot implementation across the Army, Navy, Air Force, and Marine Corps. Flexible execution, or flexicution, became one of the core themes in working with NPSP stakeholders at the command level of each military service branch and within the Office of Military Community and Family Policy (MC&FP) at the Office of the Secretary of Defense. Data collection is currently underway and will continue through 2018.

Description of NPSP: Home visitation

NPSP can refer to a suite of primary and secondary prevention programs and services that are targeted toward young and expectant families, and, more specifically, it can refer to the home visitation program that focuses exclusively on families who are at high risk of child maltreatment. The focus of the MSE is specifically on the

NPSP home visitation program. The NPSP home visitation program delivers secondary prevention parent education programming (DoDI 6500.05; DoD, 2012). Families who are expecting a child or who have children up to age 3 years (within Army, Navy, and Air Force) or

Each service branch uses a common metric to report the efficacy of NPSP to Congress each year, known as RTC (i.e., Report to Congress) 581. This metric is a ratio calculation that uses only families who completed at least six months of NPSP HV services:

Total N of Families who had no substantiated child maltreatment reports within 12 months of completing HV Services
Total N of Families that began receiving HV services in the previous fiscal year and continued for at least 6 months

5 years (Marine Corps) and who are assessed as high needs are eligible to participate in NPSP home visitation.

The screening process is consistent across the services, using the Family Needs Screener (FNS; Kantor & Straus, 1999). The FNS assesses risk by asking questions about family of origin violence/neglect, prior family of creation violence, stress, relationship discord, support, substance use, violence approval, self-esteem, and depression. A high-needs family is one whereby at least one adult respondent scores a 9 or greater on the FNS, is experiencing single parenthood (whether not in a relationship or due to a deployment), or who answers at least one of five automatic qualifier questions indicating that there are issues in the respondent's current family life. These questions are related to feeling out of control, identifying uncontrolled anger in the family system, feeling that life is sometimes not worth living, or having a partner who has been involved in a suspected or verified case of child or spousal abuse.

Each of the four services administers NPSP home visitation at their respective Active Duty installations, and the focus and implementation are similar to well-established civilian home visitation programs, such as Nurse-Family Partnership, Health Access Nurturing Development Services, and Early Head Start-Home Visitation. NPSP emphasizes establishing warm, caring, and healthy parenting relationships and working with parents to better understand and appreciate children's development across domains and over time (Blaisure, Saathoff-Wells, Pereira, MacDermid Wadsworth, & Dombro, 2015). Because this is a voluntary program, families are not required to participate. NPSP Home Visitors (HV) are Public Health Registered Nurses, Licensed Clinical Social Workers, Licensed Marriage and Family Therapists, or master's-level Child Development Specialists. Each military service has established criteria for which type of professional it employs in this position. HVs are expected to meet with families at least every other week. The range of sessions completed typically spans three to 14 visits (about 1 month to 7 months of services), although some families may stay active in the program longer and return to NPSP when new infants come into their family system through adoption or birth.

According to the 2014 Annual Report to Congress (DoD, 2014), the ratio goal for this metric is 85%. In Fiscal Year 2013, the actual ratio was 94%, which is up from 90% in 2010.

The relatively high standardization of program goals and use of both a common screener and reporting outcome metric across all four armed services indicated that NPSP was a strong candidate for developing a DoD-wide logic model and evaluation plan. These same characteristics also provided a strong foundation for exploring the feasibility of developing an MSE pilot study to test the plan.

Building the partnership through action research

The Clearinghouse for Military Family Readiness at Penn State (Clearinghouse) was tasked to deliver the following items in Phase One:

1. update an existing NPSP logic model (Kantor & Straus, 1999) so that it would be representative of current program priorities and practices across all four Services and
2. create a corresponding evaluation plan to examine both formative and summative questions.

In Phase Two, the Clearinghouse was charged with implementing and assessing a pilot evaluation using the updated logic model and evaluation plan.

The framework applied to complete the above deliverables was consistent with the principles of action research and practical participatory evaluation. Action research is a framework for collaboration between evaluators and community partners to produce applied research that addresses a specific issue facing a community or organization (Lewin, 1946; Small & Uttal, 2005). As program evaluation has evolved as a well-defined area of applied research, several types of participatory evaluation have been identified. Practical participatory evaluation pulls from action research principles to explicitly "involve stakeholders in the evaluation to improve the usefulness of the results" (Fitzpatrick, Sanders, & Worthen, 2011, p. 201). Evaluation studies often are

asked to address multiple needs and expectations. These can include questions and concerns internal to the agency, such as implementation quality, training of providers, and provider/client perceptions of whether the program meets the stated purpose. Questions and concerns can also originate external to the agency, particularly regarding continued funding and support of a program. Evidence of effectiveness, such as outcome evaluation studies, is important for internal and external stakeholders (Carman, 2009).

In an effort to produce knowledge that addresses practical concerns, evaluators provide the expertise for maintaining standards of good science, while partners' expertise related to the program is leveraged. The partners' perspectives, then, are the driving force behind what is examined (Fitzpatrick et al., 2011; Small & Uttal, 2005). In this type of partnership, evaluators and community partners possess the knowledge and insights to make research successful (Small & Uttal, 2005).

The joint stakeholders in this project included the NPSP program managers at the command level of each military service ($n = 4$), the MC&FP NPSP Manager ($n = 1$), and research and evaluation personnel (evaluators) from the Clearinghouse ($n = 4$). Each partner was encouraged to engage in co-developing the program theory (logic model) and research design, help select measures to assess program outcomes, and identify pilot sites for implementation. Action research principles encourage mutual responsiveness between evaluators and community partners, and, as such, the development of the project may take more time than a model of research that is more researcher-directed (Schutt, 2009). Over 14 months, a common program theory and research design were created. Afterward, an implementation plan was developed that could build on the existing strengths and known challenges across all four Services. A further 18 months of work addressed data management and security challenges and training needs for all affected personnel at the pilot sites. Implementation of the evaluation plan will consist of two years of data collection followed by analyses. Rolling implementation started at the first pilot site in August 2016, followed by the second site in September 2016, and the third site in October 2016. The remaining sites will begin implementation and data collection by the end of 2016, which will put the end of data collection into 2018.

In the beginning, a significant amount of time was spent early in the process to understand the characteristics of each service's implementation of NPSP Home Visitation. This time was critical as all stakeholders needed to gain a deep understanding about variations in management, staffing, curricula, resources, and client populations before trying to build a logic model and evaluation

plan that effectively encompassed the common programmatic elements and goals of all military services. In addition, these early months of information sharing helped establish effective and transparent communication pathways between the Clearinghouse and service partners. This fostered buy-in and the ability to develop an evaluation model that reflected the goals of each partner.

Key elements to identify for multisite program evaluations

As with single site evaluations, the focus and purpose of a multisite program evaluation need to be identified early. Broadly speaking, an evaluation's focus is categorized as formative or summative. Formative evaluations use activities to provide information to guide program improvement. These evaluations often seek information about the need for the program, the program's design and implementation, and its efficiency (Rossi, Lipsey, & Freeman, 2004). Formative evaluations often focus on answering questions internal to the program's agency. Summative evaluations often include varying degrees of formative information but focus more on answering the question "Does the program work in the way(s) it is intended?" These evaluations are often conducted to demonstrate whether a program successfully produces the intended outcome and make decisions about continued funding and allocation of resources. Thus, summative foci often are designed with the external stakeholders in mind.

Once the main purpose of an MSE is established, attention turns to identifying the program theory and development or update of a program's logic model. Key stakeholders from each site can provide valuable information about points of shared and unique implementation factors, participant characteristics, personnel training or qualifications, and priorities for program delivery. The development of a program theory and logic model that is inclusive of site differences and commonalities may result in a complex mapping of the program. However, this process can provide insight to the evaluator for areas of flexibility to explore in building the evaluation plan. As the program theory is built, the areas of commonality begin to provide a foundation for shared measures, outcomes, and goals across sites, and areas of difference can provide opportunities for tailoring measures, outcomes, and goals to specific sites. The inclusion of additional site-specific priorities is dependent on restrictions from the funding source, the amount of funding available, and the potential for additional funding at the site level.

Throughout the evaluation plan process, the evaluator needs to listen to the ongoing conversations of the stakeholders. Although the development and implementation

of an evaluation plan may appear to be fairly straightforward, undisclosed histories and divergent priorities could be operating in the background. Are there funding pressures driving the interest in the evaluation? Are there potential conflicts or a sense of resource inequality between the sites? Are there historical issues between or within sites that could affect an MSE? The answers to these questions can shape the process of developing a shared evaluation plan and inform later decisions about site-level training and support, data collection, analyses, and reports.

Herrell and Straw (2002) noted several details that must be addressed in order to answer two overarching questions about the appropriate use of an MSE: (a) What is the degree to which each site will be required to conform to the overall plan and (b) Are there adequate resources available to support and monitor the project across multiple sites, including accounting for site variation in data analyses? The elements that must be addressed include the following items:

- Are there questions in common asked by each site?
- Is the same intervention (program) being deployed at each site, so data could be pooled across sites?
- Are the parameters for the target populations the same across the sites? If not, what are the variations?
- Will each site conform to a common evaluation design (e.g., same measures, timeline of data collection, use and parameters of a control group)?
- How and by whom will activities be coordinated across sites?
- What training and ongoing support will be needed within and between sites?
- How will fidelity and quality assurance be monitored within and between sites?
- What strategies are planned to analyze the data across sites and within sites? How will analyses account for potential variations between sites?

Although many of the key elements described in this article are not exclusive to MSEs, they do represent a range of decision points that need to be identified and addressed early in the evaluation plan development process. MSEs are more complex to manage due to the addition of sites, whether those sites are in the same community (e.g., 2nd grade classrooms within a school district) or are geographically dispersed (e.g., Navy Reserve Units across the continental United States). Once the appropriateness of an MSE design is established, resources have to be calculated carefully to know if an MSE is feasible.

Design of the measurement model

A high level of common practice and focus was noted early in the review of NPSP, yet each military service also

developed unique pathways to accomplish common program goals. Services had identified a few distinctive NPSP priorities for their particular service member and family population. For example, differences in length of deployment and temporary duty separations and in the demographics of the young/new parent population created variation across the Services in their implementation of NPSP. Services also varied in their choices of curriculum used in NPSP.

The updated logic model reflected anticipated changes due to intervention primarily at the parent level as NPSP is a parent education program. In addition, the services agreed that changes might also be observed at the community and child levels (see Figure 1). One of the mandates from MC&FP was to organize the NPSP evaluation and updated logic model within a family strengths framework and coordinate, when possible, with resources from the Children's Bureau, FRIENDS National Resource Center for Community-Based Child Abuse Prevention. This coordination included reviewing a measure called the Protective Factors Survey (PFS; Counts, Buffington, Chang-Rios, Rasmussen, & Preacher, 2010) and using their online logic model builder to guide practical and measureable short-term, intermediate, and long-term outcomes. Multiple services also requested the review of materials and measures that focused on father engagement and coparenting in early intervention programming. The resulting updated logic model reflected practice by all four service branches; used language to emphasize a family strengths framework; and outlined potential avenues of change at the community, parent, and child levels.

To construct a DoD-wide logic model and evaluation plan that met the needs of all armed services, a two-tiered measurement model was proposed. Evaluation of goals held in common by all four services would be addressed in Tier 1 (CORE). This set of measures would be used across all services with specified data collection points. Using the same set of measures across the services allows conclusions to be drawn about the program as a whole. Although direct comparisons will not be made between the services, having the same set of measures allows for consistent statements to be made regarding the effectiveness of the program. Service-specific goals would be addressed in Tier 2 (FLEX) with a set of measures linked with the specific priorities and goals of each service. Tier 2 also included supplemental measures that are common in-home visitation programming or that one or more services was/were currently using. This two-tiered model allowed all services to share a common evaluation plan and set of measures that focused on shared outcomes of NPSP and also allowed each service to include additional measures at

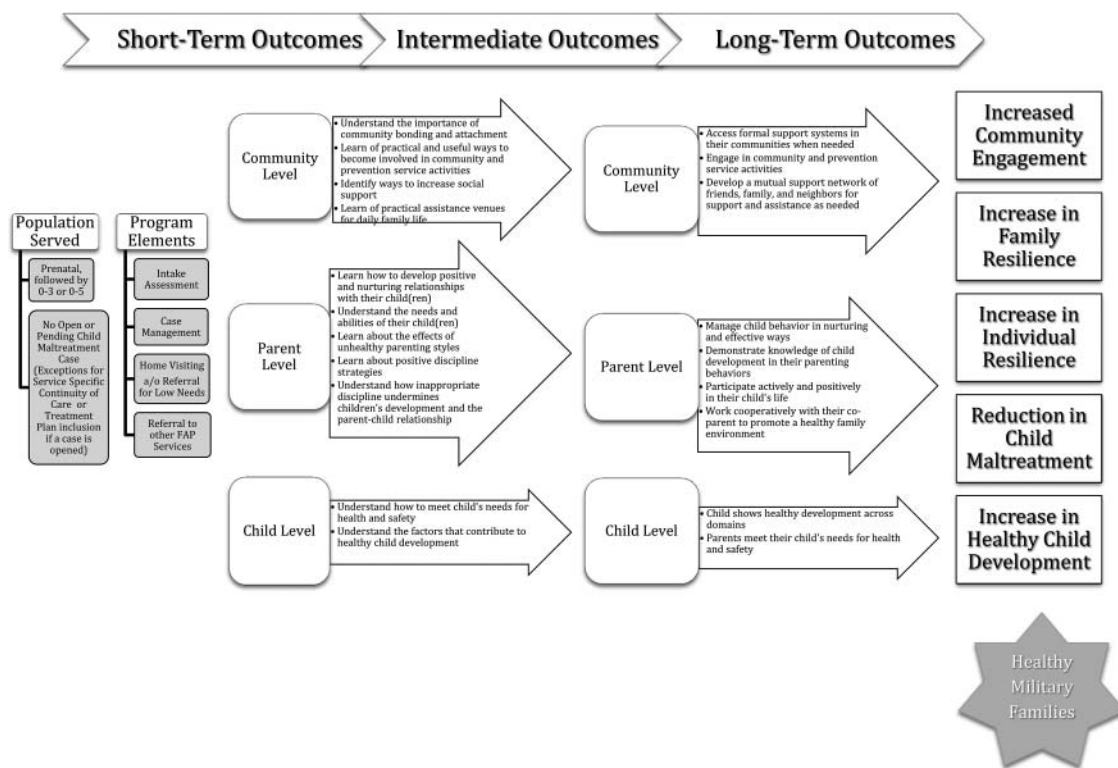


Figure 1. 2013 New Parent Support Program logic model, Department of Defense-wide.

their discretion that helped address their unique priorities. Moreover, in MSEs, some sites may feel they have to sacrifice what they want to learn about their program for what is needed by the larger organization. Having a FLEX tier may help mitigate these concerns as this approach allows each service to take ownership of the evaluation and feel their specific needs are being met.

A set of seven criteria was used to assess potential measures for inclusion in the evaluation plan. Both CORE and FLEX measures needed to meet the following criteria:

- The measure has a high level of relevance to stated NPSP goals, objectives, and expected outcomes.
- The measure has demonstrated strong reliability and validity.
- The focus of information, assessed by the measure, meets defined areas in the logic model.
- The cost of the measure for purchase and/or administration is reasonable or available in the public domain.
- The length of the measure is not prohibitive when considered as part of a data packet.
- The average time needed to complete the measure is not prohibitive to the completer.
- The level of reading difficulty is not prohibitive to the completer.

Tier 1: CORE measures

Using the four CORE measures, listed below, as part of the program evaluation strengthens the current outcome data for NPSP, particularly for the program's primary purpose as an intensive parenting education program for military families. All of the armed services currently use the FNS, and two of the services use an additional CORE measure, the Adult-Adolescent Parenting Index, 2nd ed. (AAPI-2). The current service use of one to two measures helps minimize change within the current program paperwork; potentially increases buy-in for the evaluation plan; and allows the evaluation plan to map onto current, evidence-based practices within NPSP. Keeping measurement consistent with current measure utilization is in line with the framework of action research and participatory program evaluation. Although evaluators might wish to understand everything about a program, addressing the service's needs and reducing measurement burden were top priorities. Furthermore, for the selected FLEX measures, evaluators provided choices that were consistent with the home visitation literature and relied on program stakeholders to indicate additional outcomes of interest.

Tier 1 of the measurement model efficiently enriches the depth and breadth of knowledge about how this intervention works for military families and their young children. Evaluators will be able to assess changes in participants' knowledge and utilization of

community and formal support systems, participants' stress management skills, parenting roles, parent-child interactions, parental expectations for young children, and parental use of appropriate discipline. These factors are known to affect the risk status for child maltreatment in families. The below measures represent the Tier 1 CORE and each measure's corresponding level of focus in the logic model:

- Family Needs Screener (FNS) (community and parent levels);
- Protective Factors Survey (PFS) (community and parent levels);
- Adolescent-Adult Parenting Index, 2nd Edition (AAPI-2) (parent and child levels); and
- Brief Child Abuse Potential Inventory (Brief CAPI; community, child, and parent levels).

FNS

The FNS (Kantor & Straus, 1999) specifically assesses community- and parent-level outcomes and indicators. It has been validated for use with mothers but not with fathers. However, some fathers do complete the form depending on the service and the setting in which it is administered. NPSP programs currently use the FNS across all services. It has several subscales that are appropriate for assessing changes over time, such as stress, relationship discord, support, substance abuse, violence approval, self-esteem, and depression. There are two additional subscales that focus on family of origin violence/neglect and prior family of creation violence.

PFS

The PFS (Counts et al., 2010) also specifically assesses community- and parent-level outcomes and indicators. This measure is designed as a pretest posttest survey that focuses on assessing changes in protective factors associated with reducing risk of child maltreatment and consists of four sub-scales and five single-item queries. The four subscales are family functioning/resiliency, social support, concrete support, and nurturing and attachment. The five single-item queries assess knowledge of parenting/child development, but these queries do not form a scale. Evaluators recommended that the four subscales be used for NPSP evaluation but not the single-item queries as other measures more fully assess parenting and child development knowledge.

AAPI-2

The AAPI-2 (Bavolek & Keene, 2010; Conners, White-side-Mansell, Deere, Ledet, & Edwards, 2006; Scott &

Crooks, 2007) primarily assesses parent-level outcomes and indicators but also assesses some child-level outcomes and indicators. It assesses risk for parenting behaviors, expectations, and attitudes associated with child maltreatment. Five subscales measure parental expectations of children, parental empathy towards children's needs, use of corporal punishment, parent-child family roles, and children's power and independence. Two military services currently use the AAPI-2 as part of their overall program delivery.

Brief CAPI

The Brief CAPI (Ondersma, Chaffin, Simpson, & LeBreton, 2005; Scott & Crooks, 2007) primarily assesses community-level and child-level outcomes with some focus on parent-level outcomes. It is designed to measure traits and parenting styles that are characteristic of physical child abuse and is derived from the full CAPI (Counts et al., 2010). There are three subscales comprised of 34 yes or no items (Ondersma, 2009). The Brief CAPI is comprised of the total abuse risk scale and two validity scales: the lie and random response scales.

Tier 2: FLEX measures

This second tier in the model allows for some flexibility across services. These measures include special interest areas as they align with service priorities and their respective organizational capacities to include and track data associated with these topics. The measures in this tier are grouped as FLEX measures. These requested constructs are included in a range of measures related to gauging coparenting relationships, father involvement, assessing the creation of safe and enriching home environments, and evaluating typical and atypical child development. Like the CORE measures, these measure recommendations follow the principles of action research/practical program evaluation as the constructs were specific to the requests from the individual services.

The below measures represent the Tier 2 FLEX and each measure's corresponding level of focus in the logic model:

- Nurturing Skills Competency Scale (NSCS; parent level);
- Infant/Toddler Home Observation for Measurement of the Home (HOME) Inventory (parent level);
- Massachusetts Home Safety Checklist (MHSC; child level);
- Ages and Stages Questionnaire, 3rd edition (ASQ-3; child level);
- Coparenting Relationship Scale (CRS; parent level);

- Father Engagement, items cited in Cabrera et al. (FE-Cabrera) for children under 12 months (parent level);
- Father Engagement, items cited in Leavell et al. (FE-Leavell) for children over 12 months (parent level); and
- Video-recorded 15-minute free play with mother–baby and father–baby (parent and child level).

NSCS

The NSCS (Bavolek, n.d.) assesses knowledge of parenting skills and utilization of nurturing skills and is associated with parent-level outcomes. The language used in this scale is reflective of the Nurturing Parenting Program curriculum, and the scale is intended to be used as a pretest–posttest. This scale is recommended for NPSP programs that use the Nurturing Parenting Program curriculum to help assess program and curriculum fidelity.

The HOME Inventory

The HOME (Caldwell & Bradley, 1984) assesses elements of how the parent structures the child’s daily physical and social environment. It is associated with parent-level outcomes. The HOME is an observational assessment of the following: emotional and verbal responsiveness of the parent, use/avoidance of restriction and punishment in guiding the child’s behaviors, provision of appropriate play materials and involvement with the child, and the opportunities for the child to experience variety in daily stimulation.

MHSC

The MHSC (Massachusetts Department of Public Health, Injury Prevention and Control Program, 2008) assesses elements of the child’s daily living environment and is associated with child-level outcomes. The MHSC is a tool frequently used in home visitation to assess common safety issues in different areas of the family home. Parents and HVs often do a home “tour” to identify potential risk areas and to assess emergency resources in case of an accident or injury. The areas assessed include kitchen safety, bathroom safety, child area safety, general safety, safety practices, and safety supplies.

ASQ-3

The ASQ-3 (Pizur-Barnekow, Patrick, Rhyner, Folk, & Anderson, 2010) is one of the most commonly used developmental assessments in the United States and is also associated with child-level outcomes. This assessment is commonly used throughout the NPSP program in all

armed services; however, it is often reserved for young children when the HV has determined they would benefit from early and formal identification of special developmental needs and intervention. Inclusion in the FLEX level can help program managers view this tool as useful for typically developing children and their parents, too.

CRS

The CRS (Feinberg, Brown, & Kan, 2012) is associated with parent-level outcomes and offers a systematic way to enhance understanding of the ways in which parents negotiate parenting tasks. It focuses on several aspects of coparenting: coparenting agreement and closeness, exposure to conflict, coparenting support, coparenting undermining, endorsement of partner’s parenting, and division of labor. Some of the subscales in this measure complement and further define specific elements of stress and discord in the family system, and all subscales more closely examine elements of the coparenting relationship within the larger family system.

Father engagement

There are two recommended measures for father involvement. They are associated with parent-level outcomes and assess how fathers may be engaged with their very young children. Father Engagement–Early Childhood Longitudinal Study is a set of items from the Early Childhood Longitudinal Study, cited in Cabrera, Hofferth, and Chae (2011), for fathers of children under 12 months of age. Father Engagement–Early Head Start is a set of items from the National Early Head Start Research and Evaluation Project, cited in Leavell, Tamis-Lemonda, Ruble, Zpsuls, and Cabrera (2012), and focuses on fathers whose children are 12 months through preschool age.

Video-recorded dyadic freeplay

Video-recorded free-play, separation–reunions episodes, and feeding interactions are useful to gain observational data about parent–child interactions and offer data other than parent or HV report. In early intervention research, scales such as the Nursing Child Assessment Satellite Training Parent Child Interaction and Feeding Scales are used to assess the quality of mother–infant dyadic interactions (e.g., Banerjee, & Tamis-Lemonda, 2007; Leitch, 1999) and father–infant dyadic interactions (e.g., Brophy-Herb, Gibbons, Omar, & Schiffman, 1999; McKelvey et al., 2010). Recording these sessions offers opportunities for use as a teaching tool with parents and provides opportunities for careful observational coding that may not be feasible during a live session.

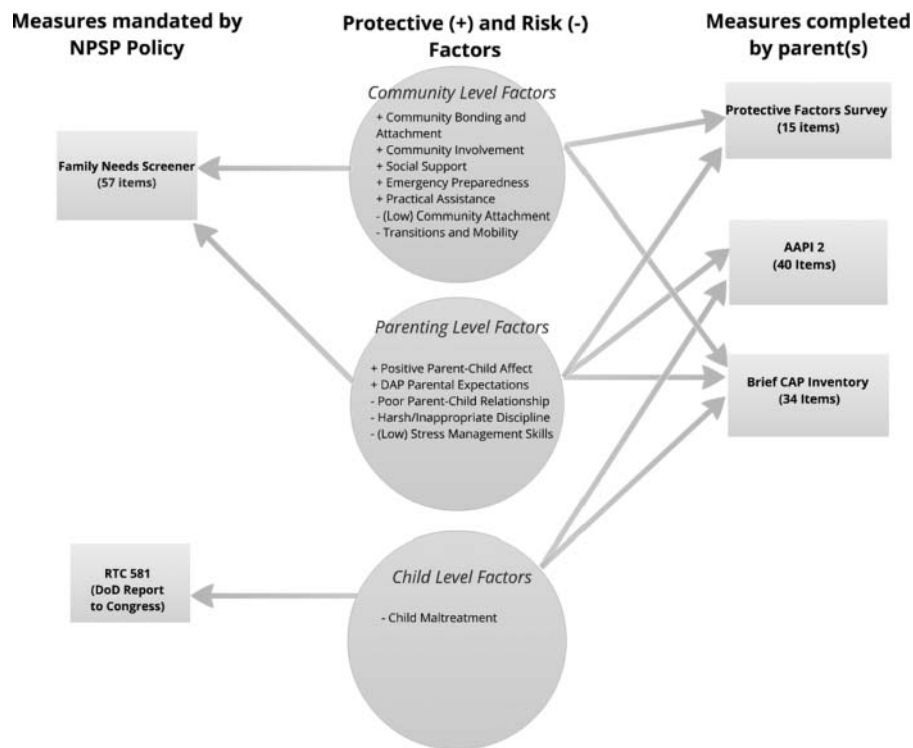


Figure 2. 2013 New Parent Support Program measure mapping by protective and risk factors: CORE measures only model.

Although evaluators knew it was not likely that the military services would be ready to test or integrate the measures in the FLEX tier of the measurement model during the pilot implementation, and, indeed, they did not, evaluators also understood that the command-level program managers were thinking long-term for the NPSP evaluation and wanted direction for assessing unique priorities. As evaluation capacity expands, the FLEX tier provides a variety of measures that can help assess additional programmatic goals. Psychometric properties and previous use in early intervention evaluation were integral to the discussions and final decisions about all measures, whether CORE or FLEX.

Stakeholder buy-in

Providing a two-tier measurement model allowed each service to be flexible in choosing additional outcomes that aligned with their unique program priorities. This also helped the services take ownership of the evaluation. Furthermore, it built structure for integrating additional outcomes as evaluation capacity strengthens over time.

Evaluators created measurement maps for the CORE and FLEX tiers in order to help stakeholders visualize the linkages between the measures and the outcomes (see Figures 2 and 3). Because the measurement model was complex, evaluators provided clear visualizations of

the measures to give the program managers a fuller understanding of what information they could expect to gain from CORE and FLEX measures. These maps also clarified the relative participant and HV burdens for completing measures and entering data.

Making choices for the common evaluation research design

Program managers, in concert with their lead data systems managers, provided important information regarding their typical clients' length of participation and potentially significant differences between civilian and military HV participation. This information was vital in creating a set of research design and comparison group options that minimized challenges and capitalized on military family characteristics. Evaluators offered a range of research designs and comparison group options, so program and data managers could make choices to fit their evaluation needs and resource capacity. After much discussion and data information gathering on NPSP participation across the military services, the common research design was comprised of NPSP participants only (i.e., no comparison group), used only the CORE measures, and contained three data collection points:

- T1: Intake,
- T2: Four months into the NPSP intervention, and
- T3: Seven months into the NPSP intervention.

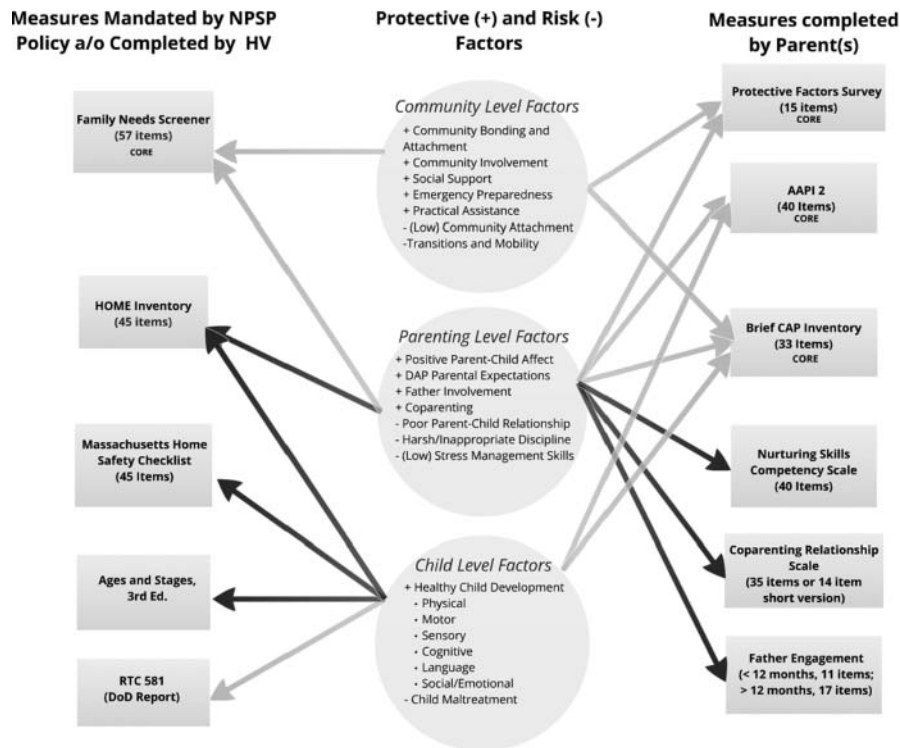


Figure 3. 2013 New Parent Support Program measure mapping by protective and risk factors: Comprehensive model.

Table 1 demonstrates the final research design and displays measures used at each data collection point. The timeline of T2 and T3 reflect what is already known about parent participation in NPSP and the minimum qualifications for families being included in the calculations for the annual metric reported to Congress, RTC 581. Services reported that their data showed participation began to drop off significantly around months five through seven. Thus, evaluators proposed a mid-intervention data collection point to gather as much information as possible about participating families before the drop off occurred. The program and data managers agreed that a 4-month time point was a satisfactory compromise to measure possible changes before loss of participation and would give insight into the population that receives less intervention, including potential for reducing risk with less intervention. Program managers were continuously concerned about dosage effects of the program. Therefore, this mid-intervention data point, when paired with data about the number of kept appointments, may help answer some questions about dosage.

The 7-month data collection point gathers information about those families who continue with the program at least one month past the minimum required for inclusion in the RTC 581 metric. Although originally proposed to match the 6-month data point, several stakeholders thought it was important to space out T2

and T3 more than 2 months, and 7 months was an acceptable compromise.

Clearinghouse evaluators originally included a potential follow-up data point at 18-months postintervention using the same CORE measures around the same time that the RTC 581 metric was calculated. However, because of the mobility of military families, logistically the follow-up was not viewed as feasible. Therefore, this data point was not included in the final research design.

Table 1. Final New Parent Support Program (NPSP) evaluation research design (no comparison group).

Measure	Time 1: baseline completed within 1st three HV	Time 2: 4 months from completion of T1 measures	Time 3: 3 months from completion of T2 measures	Time 4: 18 months from NPSP enrollment
Family Needs Screener	X		X	
Protective Factors Survey	X		X	
Adolescent-Adult Parenting Index-2	X Form A	X Form B	X Form A	
Brief Child Abuse Potential Inventory	X	X	X	
RTC 581				X

The scope of this evaluation will require significant change for program managers, HVs, and parents. Because of the number of additions and modifications to current practice, the program managers ultimately chose a research design that did not include a comparison group for the pilot implementation. Three comparison group options were offered as part of the research design. The first comparison group option was to compare program participants with families who were screened as high needs (i.e., eligible for the program) but chose not to participate. The second comparison group option included families who screened as low needs and, thus, were not eligible for the program. The third option was a combination of non-participating, high-needs families and non-eligible, low-needs families. The ultimate goal is to add a comparison group to the evaluation after the pilot is completed and stakeholders become more comfortable with the expanded evaluation.

Although evaluators engaged in a process of continuous feedback with program managers, evaluators never provided options that were not scientifically sound. The program managers' knowledge guided the choices that evaluators provided, yet program managers were ultimately responsible for the evaluation choices (i.e., research design and measures included). Establishing the scientific merit of these choices was the responsibility of the evaluation team.

Discussion

Using the principles of action research and participatory program evaluation, Clearinghouse evaluators purposefully worked to be inclusive and establish a collaborative relationship with all of the DoD and service partners. Taking time to cultivate trust and a deep understanding of each service's approach to NPSP home visitation service delivery was critical to being able to meet the goals of designing a multisite evaluation that built on common elements yet was flexible enough to accommodate service specific program priorities. This approach brought multiple stakeholders together both to contribute to and to hear multiple perspectives, priorities, and challenges of the NPSP home visitation program before trying to build a common evaluation plan. This was accomplished through a variety of approaches and behaviors. For example, evaluators learned and used language that was not particular to only one service when group meetings were held (e.g., use of service member and HV as inclusive terms) but did tailor language in one-on-one meetings (e.g., use of sailor for Navy, use of nurses for Air Force HVs). Evaluators used a combination of full group meetings and single service meetings to mitigate any potential for one service's voice to overpower others in

the development process and to better understand and address the unique needs and challenges each service expressed.

In addition, evaluators worked to offer choices, when possible and with sufficient detail, so the program and data managers could be active in discussions and in shaping the final, common plan. For example, when proposing the different measures, evaluators tried to give detailed information about potential overlap or duplication in measurement; known strengths and weaknesses of a measure, including reliability and validity (i.e., overall, with mothers, with fathers, and in high-needs parenting populations); and cost factors, including training and fair use of each measure. Evaluators also offered choices in research design, from simple/basic (e.g., pre- postsurvey design and no comparison group) to most rigorous/complex (e.g., quasi-experimental, multipoint design with comparison groups), so stakeholders could review the range of options and discuss which one(s) would likely be the best match for NPSP and most feasible for integrating into their ongoing evaluation processes.

Ultimately, the NPSP measures and research design needed to be relevant to and feasible for all four services. By first building an inclusive logic model, evaluators and program stakeholders were able to set the foundation for moving forward in designing a plan that built on current practices and strengths while expanding process and outcome data. As this phase of the project began solidifying, program and data systems managers began forming questions about next steps and meeting challenges in data management and training. With their guidance, the logistics of piloting the evaluation plan began to develop. Implementation for the pilot is underway for all four services, and the contributions of the program and data systems managers continue to be vital to the success of the project. The time taken to build the collaborative working relationships was crucial to the success evaluators are experiencing and will continue to be important as the pilot phase of the project commences.

We foresee using the two-tier research model in other evaluation projects with our partners. As the interests and needs for longer-range evaluation planning continues to grow, this type of model can help identify more immediate priorities for an evaluation, while also capturing longer-term and/or secondary priorities that may appear on a program manager's wish list and reflect anticipated changes in the client population. The research design for the pilot has some limitations as there is no comparison group and the timeline could be considered fairly short to see significant and lasting change due to intervention. However, the design reflects current practices and follows the typical program participation timeline of military families. We were able to build a

model that could pilot measures to capture the common foci across the services and do so in a research design that built upon current practices, while also acknowledging that there will be opportunities to do more as evaluation capacity and experience is built.

Funding

This material is based upon work supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture under Award no. 2010-48869-20781 (subaward grant from Purdue No. 8000047571) developed in collaboration with The Pennsylvania State University.

References

- Banerjee, P. N., & Tamis-Lemonda, C. S. (2007). Infants' persistence and mothers' teaching as predictors of toddlers' cognitive development. *Infant Behavior & Development, 30*, 479–491.
- Bavolek, S. J. (n.d.). *Nurturing Skills Competency Scale*. Retrieved from <https://www.assessingparenting.com/assessment/nscc>
- Bavolek, S. J. & Keene, R. G. (2010). *AAPI online development handbook* (2nd ed.). Asheville, NC: Family Development Resources, Inc.
- Blaisure, K. R., Saathoff-Wells, T., Pereira, A., MacDermid Wadsworth, S., & Dombro, A. (2015). *Serving military families: Theories, research, and application* (2nd ed.). New York, NY: Routledge.
- Brophy-Herb, H. E., Gibbons, C., Omar, M. A., & Schiffman, R. F. (1999). Low-income fathers and their infants: Interactions during teaching episodes. *Infant Mental Health Journal, 20*, 110–126.
- Cabrera, N. J., Hofferth, S. L., & Chae, S. (2011). Patterns and predictors of father–infant engagement across race/ethnic groups. *Early Childhood Research Quarterly, 26*, 365–375.
- Caldwell, B. & Bradley, R. (1984). *Home Observation for Measurement of the Environment (HOME)* (rev. ed.). Little Rock, AR: University of Arkansas Press.
- Carman, J. G. (2009). The accountability movement: What's wrong with this theory of change? *Nonprofit and Voluntary Sector Quarterly, 39*(2), 256–274.
- Connors, N. A., Whiteside-Mansell, L., Deere, D., Ledet, T., & Edwards, M. C. (2006). Measuring the potential for child maltreatment: The reliability and validity of the Adult Adolescent Parenting Inventory-2. *Child Abuse & Neglect, 30*, 39–53.
- Counts, J. M., Buffington, E. S., Chang-Rios, K., Rasmussen, H. N., & Preacher, K. J. (2010). The development and validation of the Protective Factors Survey: A self-report measure of protective factors against child maltreatment. *Child Abuse & Neglect, 34*, 762–772.
- Department of Defense. (2012). *Department of Defense Instruction 6400.05: New Parent Support Program*. Retrieved from <http://www.dtic.mil/whs/directives/corres/pdf/640005p.pdf>
- Department of Defense. (2014). *Annual report to the Congressional Defense Committees on the plans for the Department of Defense for the support of military family readiness: Fiscal year 2013*. Ref ID: 4-CA6A13F. Retrieved from http://www.militaryonesource.mil/footer?content_id=267470
- Feinberg, M. E., Brown, L. D., & Kan, M. L. (2012). A multi-domain self-report measure of coparenting. *Parenting: Science and Practice, 12*, 1–21.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Upper Saddle River, NJ: Pearson.
- Herrell, J. M., & Straw, R. B. (Eds.) (2002). *Conducting multiple site evaluations in real-world settings*. San Francisco, CA: Jossey-Bass.
- Kantor, G. K., & Straus, M. A. (1999). *Report on the USAF Family Needs Screener*. Durham, NH: University of New Hampshire.
- Leavell, A. S., Tamis-LeMonda, C. S., Ruble, D. N., Zpsuls, K. M., & Cabrera, N. J. (2012). African American, White and Latino fathers' activities with their sons and daughters in early childhood. *Sex Roles, 66*, 53–65.
- Leitch, D. (1999). Mother–infant interaction: Achieving synchrony. *Nursing Research, 48*, 55–58.
- Massachusetts Department of Public Health, Injury Prevention and Control Program. (2008). *Safe Home Checklist*. Retrieved from <https://massclearinghouse.ehs.state.ma.us/INJ/IP2915kit.html>
- McKelvey, L. M., Bokony, P. A., Swindle, T. M., Connors-Burrow, N. A., Schiffman, R. F., & Fitzgerald, H. E. (2010). Father teaching interactions with toddlers at risk: Associations with later child academic outcomes. *Family Science, 2*, 146–155.
- Ondersma, S. (2009). *BCAP Information Sheet* (rev.). Unpublished.
- Ondersma, S., Chaffin, M., Simpson, S., & LeBreton, J. (2005). A brief form of the Child Abuse Potential Inventory: Development and validation. *Journal of Clinical Child & Adolescent Psychology, 34*, 301–311.
- Pizur-Barnekow, K., Patrick, T., Rhyner, P., Folk, L., & Anderson, K. (2010). Early identification of developmental delays through surveillance, screening, and diagnostic evaluation. *Infants & Young Children, 23*, 323–330.
- Ross, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systemic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Schutt, R. K. (2009). *Investigating the social world: The process and practice of research* (6th ed.). Los Angeles, CA: Pine Forge Press.
- Scott, K., & Crooks, C. (2007). Preliminary evaluation of an intervention program from maltreating fathers. *Brief Treatment and Crisis Intervention, 7*, 224–238.
- Small, S. A., & Uttal, L. (2005). Action-oriented research: Strategies for engaged scholarship. *Journal of Marriage and Family, 76*(4), 936–948.