PennState

CLEARINGHOUSE
FOR MILITARY FAMILY READINESS

New Parent Support Program: Continuous Quality Improvement

Pilot Project Final Report

Tara Saathoff-Wells, Ph.D., CFLE

Nicole R. Morgan, Ph.D.

Carly Doucette, M.S.

Daniel F. Perkins, Ph.D.

June 3, 2020

Table of Contents

Executive Summary

The New Parent Support Program: Continuous Quality Improvement (NPSP CQI) pilot project was a two-phase endeavor that first designed and then tested a common, expanded program evaluation model for the NPSP home visitation program. Phase One (2012-2013) produced an updated, DoD-wide logic model and an expanded evaluation plan. Phase Two (2014-2020) included developing, implementing, and analyzing the findings of a pilot study that tested the evaluation plan. The pilot was conducted at six installations across all four Services. Variations and common ground in NPSP delivery and documentation across the Services were noted throughout Phase Two. Four measures were used in the evaluation project with two primary goals:

(1) Test each measure for its utility in program evaluation and quality improvement and for potential overlap between measures.
(2) Gauge the effectiveness of each measure for program planning from a home visitor's perspective.

Analyses support the below-noted recommendations. These recommendations have the potential to impact each Service across a range of practices including intake procedures, client assessment, and documentation practices. Careful consideration of resource allocation and programmatic change is warranted as changes may be experienced differentially.

Recommendation 1: *A comprehensive review of the Family Needs Screener (FNS) should be completed with the following aims:*

- *Identify updates that are relevant to better assess child maltreatment risks in today's military families.*
- *Refine the FNS scoring methods and eligibility criteria.*

Recommendation 2: *FNS screeners, more than 2 months old, should be reviewed by the client and home visitor to ensure updated responses are recorded in the Services' respective client management systems CMS, so that current data are driving services.*

Recommendation 3: *Analyses support using the FNS 57-item version, limitations withstanding.*

Recommendation 4: *The Protective Factors Survey (PFS) is not recommended as a pre-posttest or as a standalone measure for program improvement.*

- *However, the PFS is useful for home visitors in building rapport, program planning, and using a strengths-based approach in teaching about protective factors.*
- *A set of assessments is required in which at least one of the other measures identifies risk for child maltreatment.*

Recommendation 5: *The Brief Child Abuse Potential Inventory (BCAP) demonstrates change in risk over time and could be a useful assessment tool for NPSP when used in concert with other measures such as the FNS.*

- *The validity check procedures can help identify potential clients who might otherwise be misclassified into either low or high needs groups. In this sample, the Lie scale also demonstrated potential to highlight social desirability responses in other measures (the FNS and PFS).*
- *Note, if the BCAP is employed, training and support for home visitors and supervisors will be critical, so they can learn how to best use the measure for program planning, discussions with clients, and program improvement.*

Recommendation 6: *Continued or future use of the Adult-Adolescent Parenting Inventory, 2nd Ed. (AAPI-2) is not recommended for the NPSP program. The sunsetting of AAPI-2 involves problems with psychometric reliability, performance, and overall data management.*

Introduction

The New Parent Support Program: Continuous Quality Improvement (NPSP CQI) pilot project was a two-phase endeavor that first designed and then tested a common, expanded program evaluation model for the NPSP home visitation program. NPSP offers universal and secondary programming for new and expectant parents and their children ages 0 to 3 years (5 years old in the Marine Corps) and is available at all active duty stations. The NPSP home visitation program is a secondary prevention program that intends to prevent child maltreatment by reducing families' risk factors and strengthening families' protective factors in the early years. The home visitation program is an intensive, one-on-one program in which home visitors work with a parent (family) to set goals and share content to support positive parenting, reduce stress through social and instrumental support connections, and foster healthy relationships between parents and their children. Visits typically occur bi-weekly, and participation ranges from 1 month to more than 1 year.

In Phase One, 2012-2013, an updated NPSP logic model (Appendix A) and a subsequent CQI program evaluation plan were developed. The Clearinghouse for Military Family Readiness at Penn State (Clearinghouse) collaborated with the Office of the Secretary of Defense, Office of Military Community and Family Policy and Service NPSP Program managers.

In Phase Two, 2014-2020, a pilot study of the expanded evaluation plan was developed and implemented across six sites, with at least one site per Service. This report gives a brief summary of Phase One work completed and an in-depth report for Phase Two accomplishments and deliverables.

Phase One: Logic Model and Evaluation Plan Development

The Office of Military Family and Community Policy (MC&FP) and Family Advocacy Program (FAP) managers from each Service collaborated to select several family-based programs for expanded program evaluation support. Evaluation plans for each program were developed or updated, and established documentation for how each program was designed to affect targeted outcomes. The evaluation plan development was authorized under the guidance of the Department of Defense Instruction (DoDI) *1342.22: Military Family Readiness*. The New Parent Support Program home visitation program was one of the programs chosen, in part, because of a strong standardization in purpose and delivery across the Services, albeit with differential implementation.

During this initial phase, Clearinghouse scientists first completed a comprehensive review of existing NPSP materials that include the following: the NPSP *DoDI 6400.05* (2012), which outlines the policies, procedures, and responsibilities for the program; Service-specific operational manuals; and previously published documentation on NPSP logic model development and screening tools (Kantor & Strauss, 1999). The team also identified current promising practices, research protocols, and assessment measures within the peer-reviewed literature on home visitation programming. Program managers from each Service identified current practices regarding use of manualized curricula and measures that screened for eligibility, informed practice and planning, and measured outcomes.

This review established that there was no common, standardized use of a manualized curriculum; rather, there were at least two proprietary curricula and several parent education resources that were used when planning home visits with families. However, two measures were standardized across the Services:

1. The Family Needs Screener (Kantor & Strauss, 1999), which is completed at intake and determines initial eligibility for home visitation services.
2. The Social Compact Metric, *RTC 581* (n.d.), which is a ratio calculation of clients who completed a minimum of 6 months of NPSP services and who did not have a confirmed child maltreatment case 12 months after closure of NPSP services. This metric is reported annually to Congress.

Program managers shared information regarding Service-specific aspects of staffing and implementation and additional information that would be valuable in an expanded evaluation plan (e.g., father involvement, curriculum effects on program outcomes, implementation fidelity).

A DoD-wide logic model was developed and agreed upon by all partners. From this logic model, a two-tiered evaluation plan was proposed. The first tier was comprised of four core measures that all Services agreed to use in piloting the updated evaluation plan. The second tier included several measures that were of interest to one or more Services and, if chosen by a Service, the measure (s) could be included in the pilot for specific sites. While there was strong initial interest in the Tier 2 measures, all the Services ultimately chose to participate in the pilot with Tier 1 measures only. The general sentiment was that it was important to see how the agreed upon measures performed first before expanding to test additional measures. See Appendix B for Tier 1 and Tier 2 measures. The four measures selected for the pilot are listed here:

- Family Needs Screener (FNS, Screener)

- Protective Factors Survey (PFS)
- Brief Child Abuse Potential Inventory (BCAP)
- Adult-Adolescent Parenting Index, 2nd edition (AAPI-2)

Several research designs were considered, including designs with and without a control group. Ultimately, the final design utilized a participant-only approach (i.e. no comparison group). The design consisted of three data collection intervals, which were selected to reflect the available data from the Services for average length of participation. In addition, maltreatment data for CQI participants would be identified and pushed to the Clearinghouse at the end of the data collection period. Figure 1 illustrates the final design with intervals and measures.

Table 1
2013 NPSP CQI Evaluation Research Design (No Control Group)

Measure	**Time 1**: Baseline (completed within 1st 3 visits)	**Time 2**: 4 months from completion of T1 measures	**Time 3**: 3 months from completion of T2 measures
Family Needs Screener (FNS)	X		X
Protective Factors Survey (PFS)	X		X
Adolescent-Adult Parenting Index-2 (AAPI-2)	X Form A	X Form B	X Form A
Brief Child Abuse Potential Inventory (BCAP)	X	X	X

Note: Maltreatment data collected on participants at the end of the data collection period.

The primary purposes of the CQI pilot evaluation were as follows:

1. Test each measure for its utility in program evaluation and quality improvement and for potential overlap between measures.
2. Gauge the effectiveness of each measure for program planning from a home visitor's perspective.

Estimates for respondents were calculated based on FY2015 enrollment and 6-month retention numbers at the anticipated pilot sites. CQI participation was estimated at a 50% rate of agreement to participate and a revised to about a 40% rate of agreement, with

attrition expected to approximate the 6-month retention figures. Calculations are represented in Table 2.

Table 2
FY15 NPSP Enrollment and Retention Data from Anticipated Pilot Sites with Participation and Attrition Estimates

Service	Start FY 2015	% of Overall Pop	End (6 mo retention)	% of Pop	CQI START Original	CQI END Original	CQI START 75%	CQI END 75%
Army	1025	70%	512	70%	560	280	420	210
Navy	130	8%	65	9%	74	37	56	28
Air Force	150	10%	75	10%	83	41	62	31
Marines	163	11%	75	10%	83	42	62	31
TOTAL	1468		727		800	400	600	300

Following the approval of the research design, priorities for Phase Two were established. These initial priorities included confirming pilot sites for each Service, establishing protocols and building needed infrastructure for data management, and determining training and on-going support needs for implementation.

Phase Two: CQI Implementation

The implementation phase of the NPSP CQI project ran from 2014 through 2020. This multi-step process, which is illustrated in Figure 1, was divided into three overarching stages to describe the work completed: the pre-build, infrastructure build, and implementation stages.

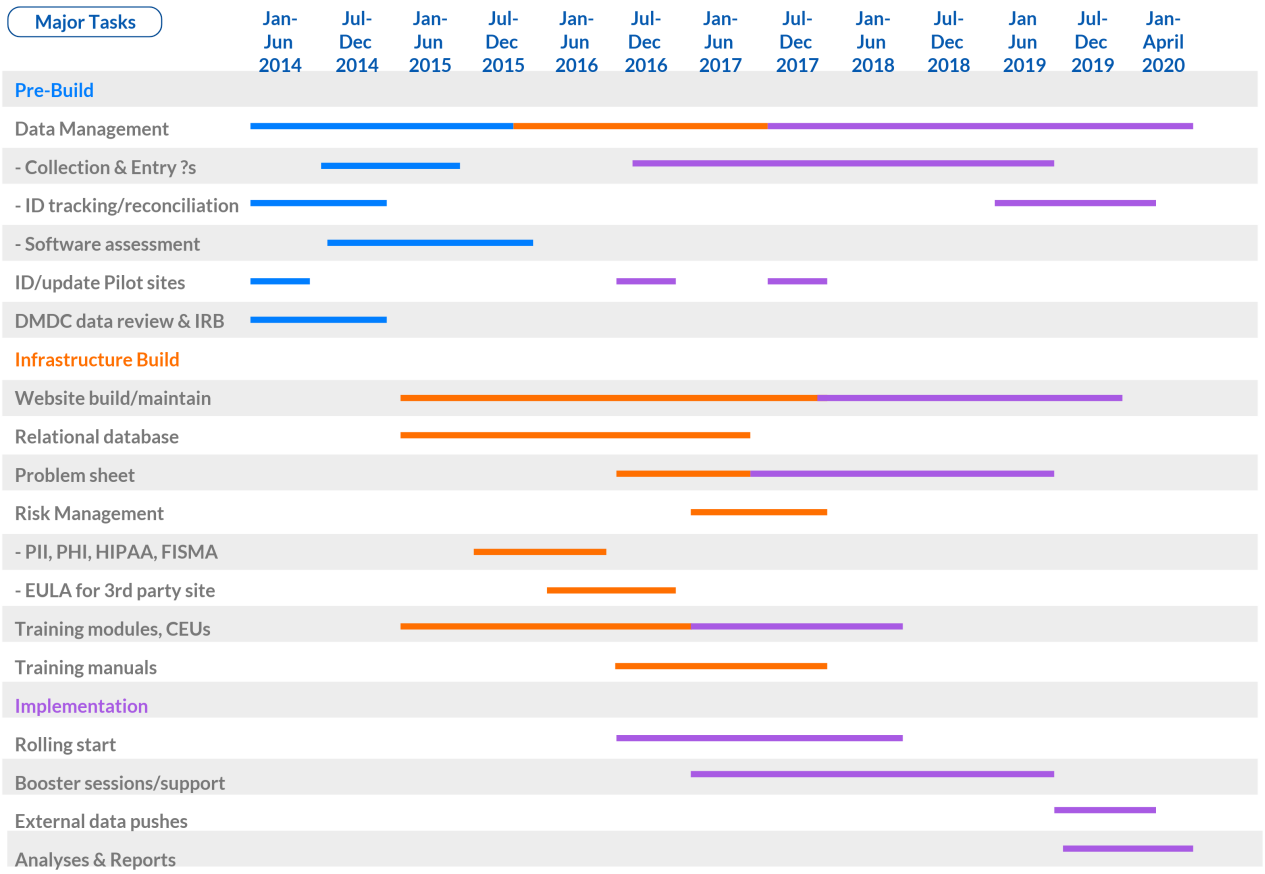


Figure 1. CQI Major Tasks Timeline, 2014-2020.

Pre-build and Infrastructure Build Stages

The pre-build stage of the project established the overall goals and priorities for Phase 2. One of the requests by the Services data analysts was that the data management system for CQI be kept as simple as possible for the sites and analysts. Thus, the data management system needed to draw upon existing Service-specific data management resources efficiently and use external systems when possible, so no changes to current client management systems would be required. This request also meant working to reduce the additional burden on home visitors as much as possible. Streamlining became

a guiding principle for all aspects of the CQI project. However, streamlining for partners meant that the Clearinghouse team assumed more complicated management roles. For example, questions about linking client data across multiple systems triggered several reviews by Penn State risk management and the institutional review board, as well as reviews by DMDC and at least one Service's research board. From these reviews, the CQI web application (app) made specific choices in data entry processes and linking unique entries into a relational database.

Pilot site identification

Each Service planned to have one pilot site each. Due to changing conditions at pilot sites during the course of Phase 2, the Navy and Air Force added one site each for a total of six sites across the Services:

Army: Fort Bragg

Navy: Naval Base Kitsap and Naval Base San Diego

Air Force: Hurlburt Air Field and Eglin Air Force Base

Marine Corps: Marine Corps Base Quantico

Points of contact included the MC&FP NPSP analyst, Service-level NPSP managers and a combination of installation-level NPSP managers and FAP managers.

Data management

The CQI data management plan needed to address two main challenges to ensure data privacy: data would come from multiple sources and would need to be linked, as seen in Table 3. Moreover, there was a range of data privacy requirements across the Services. All Services were sensitive to the use and sharing of data that contained Personally Identifiable Information (PII) and Protected Health Information (PHI), which could be subject to the Health Insurance Portability and Accountability Act (HIPAA). Data privacy and use guidelines from the Penn State Institutional Review Board (IRB) and the Defense Manpower Data Center (DMDC) also needed to be met. Table 3 lists each of the measures and their respective data systems.

Table 3

CQI Measures and Their Respective Data Management Systems

Measure	Data Entry and Scored Summary Location
Family Needs Screener (FNS)	Services' case management systems (n=4)
Protective Factors Survey (PFS)	Clearinghouse CQI Portal
Brief Child Abuse Potential Inventory (BCAP)	Clearinghouse CQI Portal
Adolescent-Adult Parenting Index-2 (AAPI-2)	Assessing Parenting system

The Clearinghouse team adopted a set of strategies to address privacy and maintain the ability to match client data across multiple systems: data de-identification and aggregated analyses. Guidance from the Privacy Technical Assistance Center (2013, archived) was integrated into CQI data management protocols:

1. Data were de-identified for CQI purposes such that PII was removed or obscured sufficiently to minimize risk of unintended disclosure of participants' identities.
2. Data had a standardized coding system so that each participant was given a unique case number that allowed for tracking participation across multiple data collection points.

De-identification of PII and PHI removes enough data such that the remaining information does not identify an individual, and there is no reasonable basis to believe the remaining information can be used to identify an individual. Two of the data entry systems could and did store PII and PHI: The Services' client management systems and the proprietary Assessing Parenting system.

The FNS contained client names in addition to the client case number, so all client names were removed from FNS data prior to being sent to the Clearinghouse. The CQI project adopted the client case numbers as the standardized coding system. Case numbers were assigned at each site as part of regular case management. These case numbers allowed client data from the Services' client management systems to be matched across the other two data systems. Only the local installations had the ability to match names with the client case numbers.

Specific guidance was required for the Assessing Parenting system. The University Risk Management Office reviewed and approved the End User License Agreement (EULA) before the Clearinghouse established an account for the project. This proprietary system is tailored to a public-sector market and is associated with a parent education curriculum. However, the AAPI-2 measure is well-known and often used independently of the curriculum. PII is asked in the demographic block of questions, and some of the demographics are used in calculating scores. The system requires answers to each demographic question in order to complete a client profile. Thus, CQI established guidelines to safeguard against inadvertently storing or sharing identifiable information on a third-party site. Guidelines included use of case numbers instead of client names; to answer *unknown* or *not applicable* to questions such as income, occupation, and history of family violence; and to approximate the client's birthdate by using only month and year.

Data privacy was further supported by planning aggregate analyses and treating the sum of pilot sites as the sample. As the CQI project was a pilot of the expanded evaluation plan, numbers from individual sites could range from about 20 to 400 participants. Two of the Services had a single pilot site, and two Services had two pilot sites each. The size of individual site samples and number of sites created the potential for client re-identification if analyses were performed at the site or Service levels.

In all, data for the CQI pilot were entered into six different systems. Data from the Services' client management systems and the proprietary system would be pushed to the Clearinghouse team by the Services to be combined with the data entered directly into the CQI portal. Thus, there needed to be a secure way to push data from the external systems to the CQI team.

Institutional approvals

The Penn State University IRB office and the DMDC reviewed the NPSP CQI proposal to assess the nature of the project, what would be asked of participants, and what data from DoD and Service-level sources was required. The IRB review concluded that the pilot evaluation was not human subjects research as defined in the Department of Health and Human Services regulations. Rather, this project was categorized as program evaluation and quality improvement. The DMDC, under guidance *of DoDI 1100.13 Survey of Personnel*, concluded that the CQI pilot did not unduly burden DoD personnel, and the data sought from the Services were appropriate.

Mechanisms of data collection and tracking

Home visitors administered the measures with CQI clients at the pre-determined intervals. Pencil and paper versions of each measure were used. Electronic options of online completion were considered - whether completed independently by clients or using a tablet supplied by the home visitor - but those options were not feasible at the time. However, this strategy is recommended as an option in future data collection work as costs for tablets have decreased and use of tablets in home visits has become more common. Electronic options could also reduce the burden placed primarily on home visitors to administer and then enter the additional measures.

Home visitors and, at a few installations, administrative support staff entered data into the systems described in Table 3. Each of the systems produced a measure printout of a client's score and interpretation of those scores. These reports were available for home visitors to use as part of visit planning.

The FNS data entered into the Services' client management programs required a data push from the site or their Services' data analysts. Data entered into the Assessing Parenting system also needed to be sent from the Services for Army, Navy, and Marine Corps. These Services were using the AAPI-2 at their pilot sites and had active accounts within the system. The Clearinghouse team set up a separate account that accommodated CQI training needs and data entry for the Air Force as this was a new measure for their pilot sites.

Secure data transfers from sites to the Clearinghouse were accomplished through use of the AMRDEC Safe Access File Exchange (2016-2018) and by its replacement DOD SAFE (2019-2020: https://safe.apps.mil/). These systems provided extra layers of encryption and protection for both sender and receiver.

The CQI portal served as the master data tracking system. This was a custom-built web app designed to provide home visitors with dashboards to track their CQI participants. Home visitors could see their whole CQI caseload on the main dashboard (Figure 2) and click on any client ID to see a specific person's progress (Figure 3). This tracking system had several features to help home visitors complete data collection at each time point.

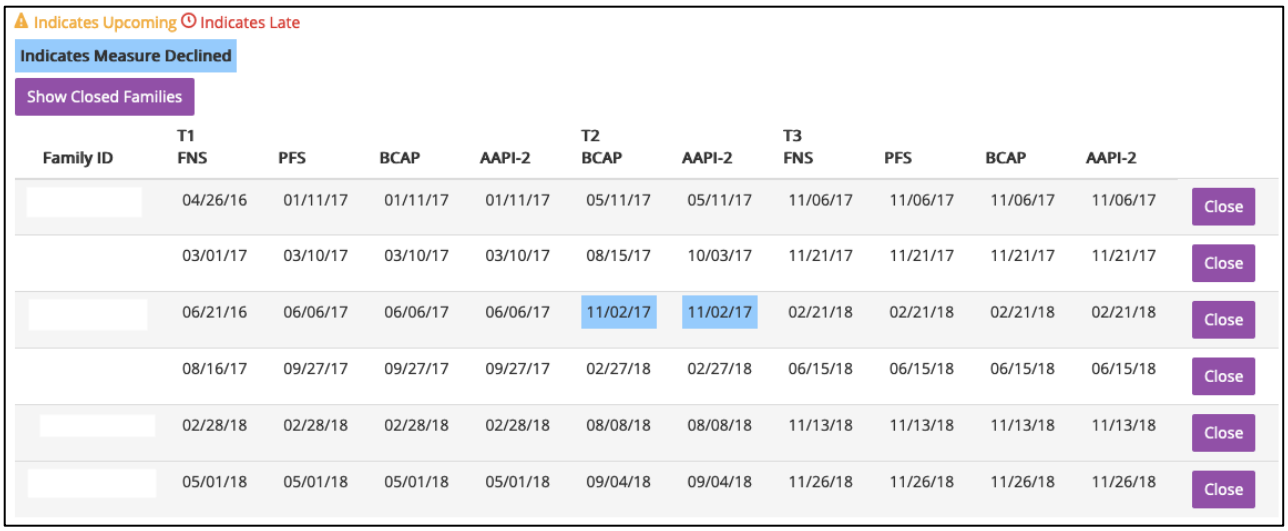


Figure 2. Home Visitor Main Dashboard.

The main dashboard listed all current participants for a home visitor, each client's status in completing the measures at all three time points, and prompts for upcoming or overdue measure completion. The home visitor could also close out a client so that client would no longer appear in the active list.

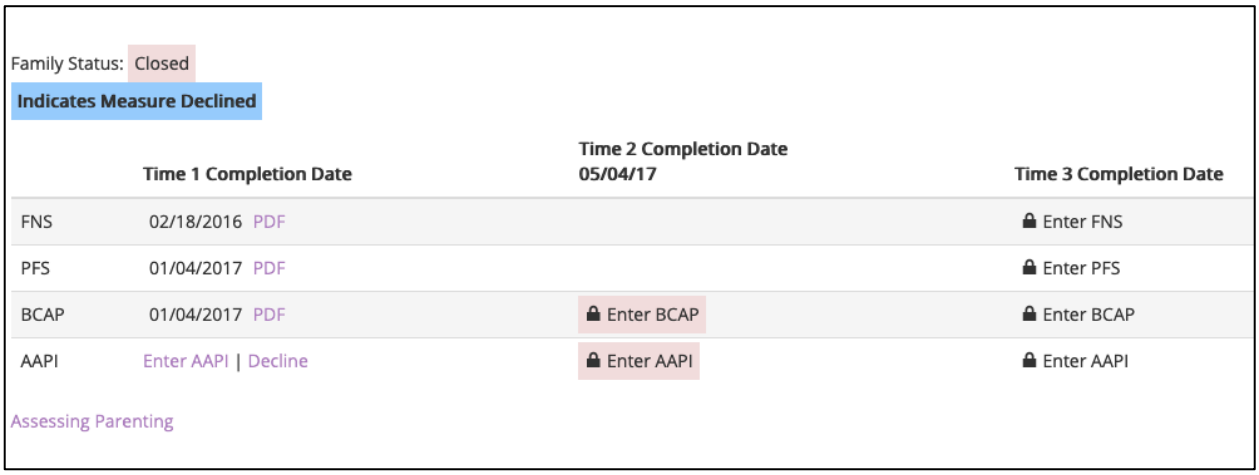


Figure 3. Client-Level Dashboard.

The client-level dashboard was designed to track each measure at each of the data collection points, and there were several features to support home visitors. Features included links to input information for each measure that, when those links were completed, showed a completion date. The Time 1 FNS and all AAPI-2 links asked for information regarding the date the client completed the measure. To reduce burden, home visitors were not required to enter those data a second time.

If a client chose not to complete a measure, the home visitor used the *Decline* link for that measure, so the system would recognize that measure as complete and generate the next data collection target date. In addition, measure links for Time 2 and Time 3 were locked, as indicated by the lock icon, until the previous measures were completed in the system. This reduced potential data entry error by limiting access to links for a single time point. The tracking and date calculation features were linked to an email reminder system,[1] so home visitors received emails twice a month with a list of all their clients who had upcoming (within 30 days) and/or overdue measures.

As seen in Figure 3, the system calculated the target date for Time 2 when all Time 1 links were completed. Likewise, when Time 2 links were completed, the system would calculate the date to administer the Time 3 measures. A direct link to the Assessing Parenting system was added to the bottom of the client-level dashboard to remove the need to open an additional web page in order to enter AAPI-2 data.

CQI web app users were assigned specific levels of access. Home visitors only had access to the clients they added to the system. Installation supervisors were able to see all of their site's home visitor caseloads. If a client changed home visitors while active in the project, the installation supervisor called the Clearinghouse, and IT support made the changes within the system.

The Clearinghouse team created a tracking system to monitor and address user challenges, including data entry errors. Thus, the data entry links were designed so a user could not go back into a completed link and change any submitted information. All data changes were made by a CQI team member and documented.

Training

A five-module training set for the CQI pilot was created that could be used in face-to-face and asynchronous settings. Two desktop manuals were printed to supplement the training: The NPSP CQI Web Registration Quick Guide and the Provider Manual. The purpose of the training was to orient NPSP personnel at each of the pilot sites to the project, to establish the goals for implementation, and to build proficiency in administering and interpreting unfamiliar measures. In a face-to-face setting, the training was held over

[1] The original email reminder system was completely automated within the app. However, it had intermittent issues in working properly and was taken offline in June 2017. From June 2017 through September 2019, a semi-automated system was employed where a member of the CQI team verified all upcoming collection dates and sent the emails.

1.5 days. Asynchronous users reported that training with phone support from the Clearinghouse team took about 2 days to complete. The trainings were available on the CQI website, which had a training version of the live site. Three case study families were constructed to provide practice opportunities with realistic situations throughout the training. No matter how the training was completed, each user had several opportunities to practice using the dashboards.

The five modules are listed below.

1. Introduction to the NPSP CQI pilot project
2. The Protective Factors Survey (PFS): Administration, scoring, and interpretation and data entry in the CQI portal
3. The Brief Child Abuse Potential Inventory (BCAP): Administration, scoring, and interpretation and data entry in the CQI portal
4. The Adult-Adolescent Parenting Index, 2nd Ed. (AAPI-2): Administration, scoring, and interpretation and data entry in the Assessing Parenting system
5. Collarelli Case Study: Independent practice with all measures

Continuing education credits were approved for social workers and nurses through the National Social Work Association (NASW) and the Pennsylvania State Nurses Association (PSNA).

Implementation

The CQI project used a rolling start, and team members conducted trainings as sites and Services were ready. The trainings occurred between August 2016 and January 2018. Two booster sessions were held, one for Fort Bragg and one for Hurlburt Air Field combined with the initial training session for Eglin Air Force Base (AFB).

- August 2016: Naval Base Kitsap
- September 2016: Fort Bragg
- October 2016: Hurlburt Air Field
- June 2017: Fort Bragg booster
- June 2017: Quantico
- November 2017: Naval Base San Diego
- January 2018: Eglin AFB and booster for Hurlburt Air Field

Installations and Services each developed preferred ways of accessing support. Some installation staff chose regular conference calls, while some home visitors preferred to

reach out individually through the Clearinghouse TA line, the problem submission form on the CQI website, or via the CQI email address npsp@psu.edu

Bimonthly updates for each site began in February 2017 and continued through June 2019, when the last enrolled participants completed their Time 3 measures. Each update was sent to the site supervisor, the Service NPSP manager, and the DoD NPSP program analyst.

Between September 2019 and February 2020, Service data personnel coordinated efforts with the Clearinghouse team to determine data transfers from each Service for their FNS and the AAPI-2 data. Air Force and Navy worked with the Clearinghouse team to set the parameters for pulling their respective FNS data sets. Trial runs of sharing worksheets with variable names were completed before data were input from their client management systems to ensure the file could be merged with existing variables. The Navy uses a shortened version of the FNS called the Family Support Survey (FSS) and has a separate form with demographic information, called the Family Social History. The Family Social History is entered into the Navy's client management system, but, currently, the FSS is not. Electronic copies of the FSS were sent via DoD SAFE and entered into the dataset by a CQI team member.

Army and Marine Corps sent electronic versions of their FNS paper forms and system printouts through DoD SAFE, and these data were then entered into the dataset by a CQI team member. Marine Corps also sent electronic copies of the AAPI-2 through DoD SAFE. Army and Navy downloaded their AAPI-2 data sets, removed non-CQI participants, and de-identified the data set before sending them to the Clearinghouse. The Air Force AAPI-2 data were available through the Clearinghouse account and were downloaded directly.

The Clearinghouse team began final data cleaning and merging waves and sites in February 2020. Analyses were conducted in March and April 2020. Analyses and recommendations follow.

Evaluation Design

As described in the Phase One summary, the NPSP CQI pilot evaluation utilized a participant-only, longitudinal design with three data collection intervals. The intervals were set to capture all participants at intake (Time 1), at about 4 months (Time 2) into NPSP services, and at about 7 months (Time 3) into NPSP services. Data from the field strongly suggested that there was a significant and common drop in participation around 4.5

months; thus, the Time 2 interval was set to assess possible changes for those who completed less than 6 months of programming. The Time 3 interval was set to gather information about clients who participated longer than the minimum time required to calculate the RTC 581 metric.

Participants

There were 243 participants across six installations (Table 4). Most participants were married (85%) but had been living together less than 1 year (77%). One third were pregnant or in the process of adopting (33%), and 46% had adopted or had had a baby within the previous 12 months. Twenty-two percent reported being pregnant but did not have a child 3 years or younger (5 years for Marine Corps), and 8% were currently pregnant and had a child who was NPSP-eligible. For families who entered into services with NPSP-eligible children, their children's age at entry into services was approximated from Time 1 PFS questions that asked for birth month and year (n=173). The average age was 7.9 months (*SD* 11.65). CQI participants entered into NPSP services between September 2016 and January 2019. Across participants, NPSP-eligible children were born between 2014 and 2018, and the majority (61%) were born in 2017.

Table 4
Participants x Service

Service	Number of sites	Primary Caregivers
Army	1	n=114
Navy	2	n=61
Air Force	2	n=49
Marine Corps	1	n=19

Active duty Service members comprised 26% of participants. The average age of clients was 27 (*r* = 18-43) and of clients' partners was 30 (*r* = 20-52). The FNS does not ask gender; therefore, gender identification data were pulled from completed AAPI-2 assessments (n = 153). Ninety-five percent (n = 145) of these clients were female. CQI participant demographics are in Table 5. Due to slight variations in demographic questions across the FNS/FSS, the number of those counted will vary across items and are in the table notes.

Table 5

Description of CQI Participants and Children, Total Sample (n = 243)

Female (n=145)[1]	**95%**
Marital Status (n=239)	
Single	11%
Married	85%
Divorced, Separated, Widowed	3%
Military Status (n=232)	
Active Duty	26%
Spouse	69%
Other	5%
Living Situation (n=176)[2]	
Living together with your partner/ spouse	85%
Living alone	9%
Other	6%
How long living together (n=176) [2]	
Less than 1 year	77%
Currently pregnant or in the process of adoption (n=177) [2]	33%
Have or adopt a baby over the last 12 months (n=178) [2]	46%
How many children are living with you? (n=166) [2]	
None	28%
1	46%
2	15%
3	10%
4	2%
Do you have any children living with you who are from a prior relationship? (n=176)	8%
Is your spouse on deployment (n=32)[3]	22%
Is the father of the baby expected to be deployed during pregnancy or for birth?	27%

Female (n=145)[1]	**Client (n=232)**	**Your partner (n=217)**
What is your/ your partner's age	Avg = 27 (R = 18 to 43)	Avg = 30 (R = 20 to 52)
What is the last year of school completed	**Client (n=232)**[4]	**Your partner (n=169)**
Some high school/ GED	5%	2%
High school graduate	18%	21%
Some College	35%	41%
College graduate	28%	24%
Post BA Training/ Advanced degree	14%	12%
Race/ Ethnicity	**Client (n=224)[4]**	**Your Partner (n=152)**
White Non-Hispanic	50%	49%
Black Non-Hispanic	25%	28%
Latino	15%	13%
Other/ Multi-racial	10%	10%

[1] Note only 153 clients had gender in the T1AAPI as it is not asked in the Family Needs Screener.

[2] Not included in Navy's Family Needs Screener.

[3] Only asked in the Navy Family Support Survey- slight wording differences between the two forms.

[4] Recode AAPI content for (n=50) Navy clients who did not have education (n=78) and who did not have Race/ Ethnicity in FNS. AAPI does not ask partner education or race/ ethnicity.

Descriptive data were available for reasons participants (n=126) dropped out of the CQI project. The most common reason given was that clients were no longer participating in NPSP services followed by completing services. Other reasons included separation from the military, permanent change of station to a new duty station or moving due to deployment. Figure 4 illustrates the drop reasons that home visitors recorded in the CQI system.

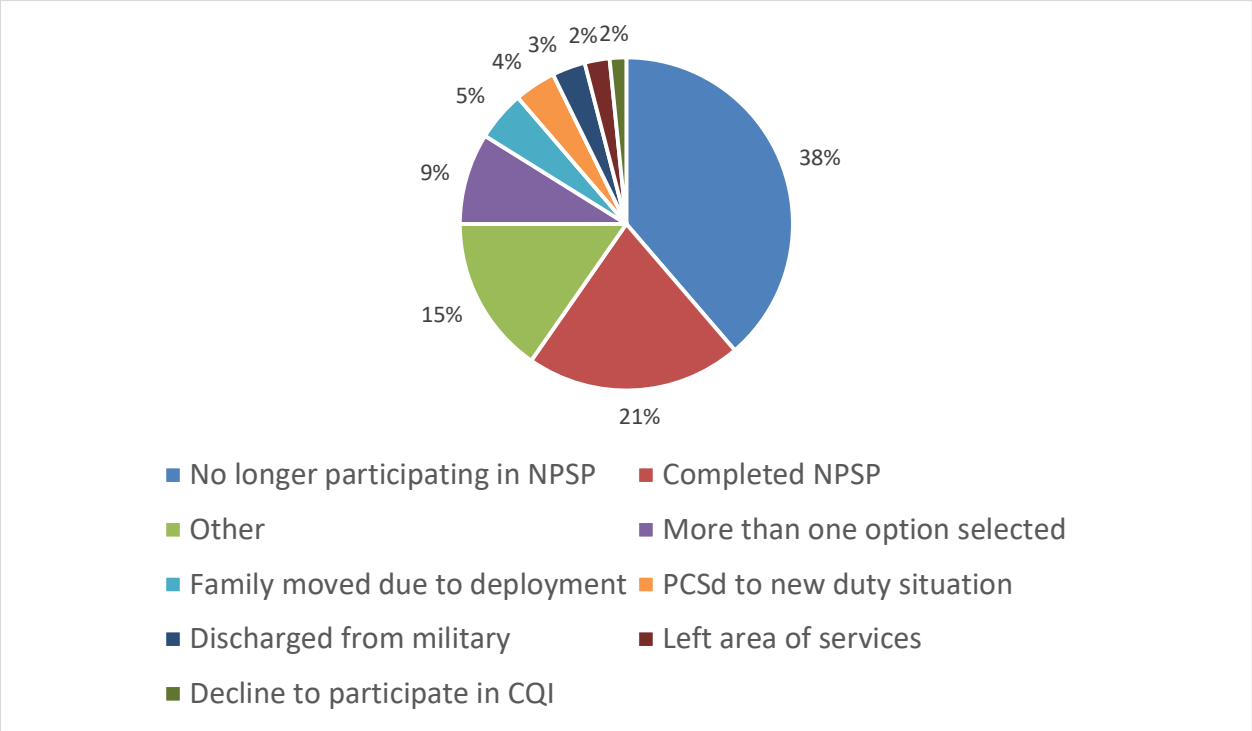


Figure 4. Drop Reasons Recorded for CQI Closure.

Procedures

Families entering into NPSP home visitation services were asked specifically to participate in the NPSP CQI project. The home visitor shared that the project's aim was to help with program improvement. The first few visits typically include administrative and rapport-building tasks, including giving signed consent to participate in the home visitation program, reviewing a client's completed FNS and any other assessments they may have completed, and establishing goals for the duration of services. A CQI informed-consent form was added to the intake process.

Families who agreed to participate completed the remaining Time 1 measures over the next three to five home visits. Home visitors were encouraged to use the CQI measures

as part of their rapport-building, whether engaging in conversations with parents about a measure or sharing feedback on scores. Families who stayed active in NPSP then completed two measures at Time 2 and repeated all four measures at Time 3. Table 1 from page 7 illustrates the data collection timeline.

Measures

Family Needs Screener (FNS) and the Family Support Survey (FSS)

Two versions of an established screener (Kantor & Strauss, 1999) were used to assess a client's eligibility for NPSP home visitation services. This screener and the home visitor's professional assessment comprised the initial evaluation for whether the intensive home visitation program was appropriate for a family. Both versions of the screener elicited information about a client's background, beliefs, and well-being across a range of risks and was purported to be sensitive to change over time. Information about FNS is presented first, and, then, information about the FSS follows.

FNS (57-items)

The Army, Air Force, and Marine Corps use the 57-item FNS to assess a client's eligibility for home visitation services. There are 13 demographic questions followed by 46 items distributed across nine subscales. Forty-four of these items are answered on a 4-point Likert-type scale that ranges from 1 = strongly disagree to 4 = strongly agree, and two items have yes/no response options.

The eight subscales that use the 4-point Likert-type scale are outlined below:

Stress: Five items that measure current perceived stress levels and pregnancy-related stress (e.g., "This is a very stressful time for me right now." "This is not a good time for me to have a baby.").

Relationship Discord: Five items that assess current distress or dissatisfaction within the client's relationship with his or her spouse/partner (e.g., "I wish my partner and I got along better.")

Support: Ten items that identify both concrete and social support. (e.g., "My income is often inadequate for basic needs [rent, food, clothing, transportation, etc.]." "There is someone I can talk to openly about anything.").

Substance Abuse: Three items that measure both the client's and their partner's alcohol use (e.g., "I sometimes drink enough to feel really high or drunk." "My partner sometimes drinks five or more drinks at a time, but mostly on weekends.").

Violence Approval: Four items that assess the client's attitude towards spousal violence and corporal punishment (e.g., "I can think of a situation where I would approve of a wife slapping a husband's face." "It is sometimes necessary to discipline a child with a good, hard spanking.").

Family of Origin Violence & Neglect: Six items that assess history of physical violence, neglect, and unhappiness within a client's family of origin (e.g., "When I was a teenager, I was hit a lot by my mother or father." "I have unhappy memories of my childhood.").

Self-Esteem: Five Items that assess the client's beliefs about his or her worthiness (e.g., "I frequently feel as if I am not as good as others.").

Depression: Four items that measure feelings of sadness and hopelessness (e.g., "There are times when I feel life is not worth living." "I think good things will happen to me in the future.").

The ninth subscale consists of two items with yes/no response options.

Prior Family Violence: Two items designed to identify occurrence(s) of prior family violence - either child maltreatment or spousal violence (e.g., "Have you or your partner been involved in suspected or verified… child abuse or neglect?" "…spouse abuse or neglect?").

FNS Scoring

Sum scores are calculated for the nine subscales and then added together for a total sum score of all forty-six items. These items are a combination of positive and negative statements. In order to calculate the sum scores, teach item needed to be recoded. The recoding scheme transforms the original 1-4 scale into 0 or 1. Zero indicates a lack of risk or the presence of strength for an item. A "1" indicates the presence of risk. Two examples of the 0, 1 transformations are illustrated below. Fifteen items were worded in the affirmative, such as number 36 in Table 6, and as such were reverse coded from the original scale. Green indicates a lack of risk or presence of strength, while red indicates a presence of risk.

Table 6
FNS Score Conversion Examples

Item #	Strongly Disagree	Disagree	Agree	Strongly Agree
Original response choices:	1	2	3	4
Item 35. When I was growing up, I saw my mother or father hit or throw something at their partner.	0	0	1	1
Item 36. My parents helped me when I had problems.	1	1	0	0

The total sum score is used to determine eligibility, and a cutoff is used to separate high from low needs. A score equal to or greater than 9 is considered high needs.

In addition to the calculated subscale and overall scores, five items are classified as automatic qualifiers for home visitation services. If any one of these item responses indicates risk, the client is automatically qualified for services. The items include two from the Prior Family Violence scale and one each from the Stress, Relationship Discord, and Depression subscales. The Marine Corps has added questions about deployment, post-traumatic stress disorder (PTSD), and traumatic brain injury (TBI) to their FNS as part of the intake process, but they are scored only as automatic qualifiers. Thus, high needs and eligibility for home visitation services are evaluated by a sum score ≥ 9 or the affirmation of risk on at least one of five specific items[2].

FSS (25-items) and Scoring

The FSS is a shortened version of the FNS and is used across the Navy. There are 25 items drawn from the FNS and the same 4-point Likert type scale, which ranges from 1 = strongly disagree to 4 = strongly agree, is used. The FSS has between 9 and 11 demographic questions, dependent on installation variations in the form.

Similar to the FNS, a sum score is calculated and a score of 4 or greater qualifies as high needs. There are no subscales to score. The FSS uses the same set of automatic qualifier

[2] Across the Services, home visitors regularly exercise clinical judgement in addition to reviewing the FNS/FSS scores. If a client's score does not meet the threshold of high needs, but there are observable signs of needs, the home visitor can override the score and offer home visitation services to the client.

items. If any one of these items is answered affirmatively, it receives a score of 4, which also meets the cutoff threshold for services.

Protective Factors Survey (PFS)

The PFS (FRIENDS, 2011) is a 20-item pre-post measure that is designed to be used in programs that provide child maltreatment prevention services. The PFS utilizes a family strengths framework that engages parents by identifying areas of strength and resilience. It measures five domains of protective factors:

1. Family functioning and resiliency
2. Social support
3. Concrete support
4. Nurturing and attachment
5. Knowledge of parenting and child development.

For the purposes of the CQI project, the domain of knowledge of parenting and child development was not assessed through this measure as another measure also assessed this construct.

Family Functioning and Resiliency: A five-item scale that measures adaptive skills and strategies for perseverance in times of distress and crisis. This scale captures a family's ability to communicate openly about positive and negative experiences to facilitate problem-solving (e.g., "In my family, we talk about problems.").

Social Support: A three-item scale that assesses a family's informal emotional supports from friends and family (e.g., "I have others who will listen when I need to talk about my problems.").

Concrete Support: A three-item scale that measures access to tangible supports (i.e. goods and/or services) that help families cope with crisis or stressful times (e.g., "I wouldn't know where to go for help if I had trouble making ends meet.").

Nurturing and Attachment: A four-item scale that assess the emotional connection and positive interactions between the parent and child that grow over time (e.g., "I am able to soothe my child when he/she is upset.").

PFS Scoring

Two 7-point Likert-type scales provide response choices in the PFS (Table 7). The first set of responses ranges from 1 = Never to 7 = Always. The second set of responses ranges from 1 = Strongly Disagree to 7 = Strongly Agree.

Table 7
PFS Response Choices

Never	Very Rarely	Rarely	About Half the Time	Frequently	Very Frequently	Always
1	2	3	4	5	6	7
Strongly Disagree	Mostly Disagree	Slightly Disagree	Neutral	Slightly Agree	Mostly Agree	Strongly Agree

The Family Functioning and Resiliency and Nurturing and Attachment scales use the Never to Always response choices, while the Social Support and Concrete Support scales use the Strongly Disagree to Strongly Agree choices. The three Concrete Support items are reverse coded as they are worded in the negative. All other items are worded positively. Average scores are calculated for each scale. First, all responses are added and then divided by the number of items in that scale to find the average, which can range from 0 to 7.

A color-coded system was used to assist in score interpretation for the PFS (Table 8). Low mean scores on the PFS are associated with low levels of protection from the measured protective factors and higher levels of risk. Conversely, high mean scores on the PFS are associated with high levels of protection from the measured protective factors and lower levels of risk.

Table 8
PFS Score Coding System

Mean Score	Color Code	Level of Protection	Category of Risk
0-3	Red	Low	High
4	Yellow	Neutral	Moderate
5-7	Green	High	Low

Brief Child Abuse Potential Inventory (BCAP)

The BCAP (Ondersma, Chaffin, Simpson, & LeBreton, 2005) is a 33-item, shortened form of the CAP Inventory (CAPI) that is designed to assist individuals who work in child protective services during investigations of reported child abuse. The BCAP is a screening tool that measures traits and parenting styles that are characteristic of *physical* child abusers. If an individual's scores are in the at-risk range, the score results do not mean he or she is maltreating his or her child; instead, further assessment is required just as if the full CAPI had been used.

The BCAP is comprised of three scales, one that assesses risk and two that provide validity information about a respondent's risk score.

Abuse Risk: This is the primary scale and contains 24 items taken directly from the full CAPI Abuse Risk scale. The Abuse Risk scale is the primary score used for interpretation. As with the full CAPI, the BCAP Abuse Risk scale is derived from items that are related to happiness, loneliness, family conflict, rigidity, distress, poverty, and feelings of persecution (e.g., "I am a happy person [happiness]." "Everything in a home should always be in its place [rigidity].").

Lie: This scale is comprised of six items designed to identify responders who may be answering items dishonestly (e.g., "Sometimes I have bad thoughts."). If four or more of these items are endorsed, the Abuse Scale should be interpreted with caution.

Random Response: This scale is comprised of three items that are written in a way that seems illogical (e.g., "Children should not learn how to swim."). This helps identify responders who are simply filling out answers instead of carefully reading the items and selecting the most appropriate responses. If any of these items are endorsed, the Abuse Scale should be interpreted with caution.

BCAP Scoring

The BCAP uses Agree/Disagree response options for each item. The Agree and Disagree responses are summed separately; 3 of the 33 items earn scores of 1 if the respondent chooses Disagree (Items 1, 22, 28). The remaining 30 items earn a score of 1 if the respondent chooses Agree. The number of Agrees and Disagrees are summed to produce an abuse risk score, which can range from 0 to 24. Two cutoff scores are used with this brief version, and these are associated with the lower and upper cutoffs used in the full version. The lower BCAP cutoff is 9 and the upper cutoff is 12. Similar to the PFS, a color-coded system was provided to help home visitors interpret the Abuse Risk Score Table 9.

Table 9
BCAP score coding system

Risk Abuse Scale Score	Color Code	Category of Risk
0-8	Green	Low
9-11	Yellow	Moderate
12+	Red	High

The Lie scale can range from 0 to 6, and, if a respondent scores 4 or greater, the Abuse Risk score may not be valid. The Random Response scale score can range from 0 to 3. Any score greater than 0 may indicate that the Abuse Risk score is not valid.

Adult-Adolescent Parenting Inventory (AAPI-2)

The AAPI-2 (Bavolek & Keene, 2001) is a 40-item, norm-referenced measure that assesses high-risk parenting attitudes and behaviors. It was last re-normed in the late 1990s. The AAPI-2 is designed to assist professionals in assessing the parenting and child-rearing attitudes of adolescent and adult populations. Developed from the known parenting and child-rearing practices of abusive and neglecting parents, AAPI-2 responses indicate degrees of agreement and disagreement with maladaptive parenting behaviors. As such, responses on the AAPI-2 provide an index of risk (i.e., high, medium, or low) for practicing abusive and neglectful parenting and child-rearing behaviors. The AAPI-2 has two versions, Form A and Form B, and items differ in each of the forms[3]. These two forms are used in an alternating format when the measure is used two or more times.

The AAPI-2 has five scales:

Appropriate Expectations: Maltreating parents tend to inaccurately perceive the needs, skills, and abilities of their children. When parents hold inappropriate

[3] There is concern that Forms A and B are not equivalent. The most common way to develop multiple forms is to reorder the same items from one to the other. The AAPI-2 Form B uses several unique items not found on Form A and more negatively worded items. In the manual, one item that is on both versions is in Construct E – Children's Power and Independence on Form A and Construct B – Empathy on Form B: "Parents who encourage their children to talk to them only end up listening to complaints."

expectations for their child's age-appropriate behaviors, there is an increased likelihood of inappropriate responses to their child (e.g., "Parents spoil babies by picking them up when they cry.").

Empathy: Parents who have low empathy are at higher risk of parental aggression toward their children; however, this has been found to differ between mothers and fathers (e.g., "Parents who encourage their children to talk to them only end up listening to complaints.").

Corporal Punishment: Many parents have strong beliefs in the use and value of physical punishment, while others voice strong beliefs against physical punishment. In addition, a substantial parent population is reluctant to use physical punishment but does not know how to use other, non-violent strategies effectively and may be ambivalent about using corporal punishment (e.g., "If a child is old enough to defy a parent, then he or she is old enough to be spanked.").

Family Roles: Parent-child role reversal is a risk factor for maltreatment. Role-reversal relates to inappropriate expectations such as children being responsible for their parents' happiness (e.g., "Children should be their parents' best friend.").

Power and Independence: Abusive parents are more likely to hold strong beliefs about the need to control children's behaviors, so children will be obedient to their authority. At the other extreme, neglectful parents may not set limits that ensure safety as children explore and expand their skills and abilities (e.g., "Children need to be allowed freedom to explore their world in safety.")."

AAPI-2 Scoring

As shown in Table 10, this measure uses a unique, 5-point Likert scale for all items. This response set was designed to try to reduce random responding.

Table 10
AAPI-2 Response Options

Strongly Agree	Agree	Disagree	Strongly Disagree	Uncertain
1	2	4	5	3

Data for this measure must be entered into the Assessing Parenting data system, which then transforms raw scale scores into standard ten (sten) scores ranging from 1 to 10.

- Sten scores between **1 and 3** indicate a high risk for abusive parenting behaviors.
- Sten Scores between **4 and 7** represent a moderate risk for abuse.
- Sten scores between **8 and 10** indicate positive, nurturing parenting attitudes and a low risk for abuse.

The creators indicate that gender and age of respondent are weighted variables in analyses such that a raw score of 28 for one person could equal a sten score of 6, while the same raw score for a person with different demographic characteristics could have a sten score of 8. This means that the first person is in the moderate risk category while the second person is in the low risk category. At the time of writing this report, the Clearinghouse team is awaiting responses to several questions about the weighted variables and calculations from raw to sten score.

Analyses

Analyses were conducted to identify and understand potential timing variations across the Services in measure completion at Time 1. Anecdotal reports about common practices indicated that there could be wide variations in how far in advance an FNS/FSS might be completed before a client entered into home visitation services. In addition, there was variation in how often the FNS/FSS was used and if a Service's client management system could accommodate multiple FNS entries without overriding a previous entry.

Analyses were also conducted on measure reliability, correlations of measures to each other, predictive abilities where appropriate, and sensitivity to change over time. The longitudinal analyses are modest due to the attrition at Times 2 and 3.

Measure Administration at Intake (Time 1)

A feature of the CQI data entry system included each measure's completion date being entered at each time point. The primary purpose of this was to help establish the Time 2 and Time 3 dates for data collection; nevertheless, this feature also supported straightforward analyses to examine whether there was variation in FNS administration and what the time span was for completing the initial four CQI measures.

The FNS/FSS completion date compared to the other CQI measures varied from about 1 to 4 months. In contrast, the PFS, BCAP, and AAPI-2 measures were completed in a window of 0 to 10 days. Table 11 shows the average time differences by Service.

Table 11

Average Time (in days) Between Services' Administration of CQI Measures

	Army	Navy	Air Force	Marines	Results
T1 FNS – AAPI	30.66 (n=83)	97.30 (n=59)	108.60 (n=43)	42.30 (n=15)	*F*(3, 196)=9.01, p <.001
T1 FNS - BCAP	44.42 (n=89)	118.50 (n=63)	113.40 (n=47)	41.60 (n=16)	*F*(3, 196)=9.78, p <.001
T1 FNS - PFS	28.60 (n=102)	106.50 (n=63)	111.24 (n=46)	41.60 (n=16)	F(3, 223)=14.28, p <.001
T1 AAPI- T1 BCAP	-9.96 (n=74)	-0.32 (n=53)	0.00 (n=44)	-2.50 (n=16)	F(3, 182) = 0.64 p = .59
T1 AAPI - T1 PFS	3.60 (n=82)	0.26 (n=54)	0.20 (n=44)	-2.44 (n=16)	F(3, 192) = 0.26 p = .85

Thus, the completion window for the non-FNS/FSS measures is much shorter. The time between completion of the intake screener to when services started and the client enrolled in the CQI project ranged from approximately 1 to 4 months. The time differences could impact the reliability and stability of a client's protective factors such that responses given on the intake FNS/FSS could change by the time services begin.

Measure Psychometrics

Each measure was evaluated for reliability, which is the likelihood of a measure to produce similar results under consistent conditions. Cronbach's alpha provided information about how well each item in the scale/subscale measured the intended construct, which is also known as the internal consistency of a scale or subscale. Item level scores range from 0 to 1. An alpha of .70 and higher is desirable and indicates the item contributes to the identified construct (e.g. family functioning). Each item then contributes to a scale-level alpha, which has the same threshold for strength of the scale, .70 and higher. This type of analysis can also identify if a scale/subscale could be stronger by eliminating one or more items. Each CQI measure's reliability, assessed with Time 1 data, is discussed below.

Family Needs Screener and Family Support Survey

Cronbach's alpha analyses for the two versions of the NPSP intake screener were completed using the full sample for each version (Table 12). In the 57-item FNS, alphas were calculated for nine scales. The Stress subscale was split into the three items that were answered "only if pregnant" and the remaining two general stress items. Cronbach's alphas were then completed on the whole sample using only the items in Navy's 25-item version.

The 57-item FNS had higher internal consistency than the 25-item FSS in this sample, and Cronbach's alphas exceeded the .70 threshold on four of five comparable scales. Comparison was not possible for pregnancy stress or violence approval as the shorter version had only single items representing those scales. The stress scales did not meet the Cronbach's alpha threshold of .70 on either version of the screener, which indicates these items may not be measuring stress in a consistent or coherent way.

The 25-item FSS had lower internal consistency than the full FNS, well under the .70 threshold, across three scales: Relationship Discord (α = .56), Support (α = .55), and Family of Origin Violence (α = .66). Two scales were unchanged from the 57-item version to the 25-item version: the 2-item Stress (α = .51) and Substance Abuse scales (α = .64). The Support scale's Cronbach's alpha improved across the two versions by dropping the same item: "My income is often inadequate for basic needs (rent, food, clothing, transportation, etc.)." However, the FNS version, with 10 items, remained the stronger scale with a change from α = .82 to α = .84 versus the FSS 4-item version, which improved from α = .55 to α = .66.

Table 12

Cronbach's Alphas for the FNS and FSS, Full Sample

57-item FNS	Alpha	If item Deleted		25-item FSS	Alpha	If item Deleted
FNS- Stress (pregnant) (n=56)	0.62			FNS- Single item Q14 for pregnant stress	-	
FNS- Stress (2 items) (n=237)	0.51			*Same 2 items - no change*	*0.51*	
FNS- Relationship Discord (5 items) (n=160)	0.74	.78 if Q23		FNS- Relationship Discord (3 items) (n=223)	0.56	
FNS- Support (10 items) (n=170)	0.82	.84 if Q39		FNS- Support (4 items) (n=231)	0.55	.66 if Q39
FNS- Substance Abuse (3 items) (n=166)	0.64	.75 if Q28		*Same 3 items - no change*	*0.64*	
FNS- Violence Approval (4 items) (n=177)	0.68	.71 if Q29		FNS- Only single item 30 (slapping husband)	-	
FNS- Family of Origin Violence (6 items) (n=174)	0.83			FNS- Family of Origin Violence (3 items) (n=240)	0.66	
FNS- Self Esteem (5 items) (n=177)	0.82			FNS- Self Esteem (3 items) (n=236)	0.85	
FNS- Depression (4 items) (n=174)	0.84			FNS- Depression (3 items) (n=235)	0.72	

Protective Factors Survey

Cronbach's alpha analyses indicated that all four scales of the PFS had moderate to robust consistency (Table 13). The alphas suggest that this measure would consistently provide similar results under stable conditions. Two scales improved by dropping one item each. The Nurturing and Attachment scale was not answered by all respondents. That scale required meeting the condition of having an NPSP-eligible child currently in the family system. Expectant parents with no other children did not answer those items. The PFS performed similarly to the validation study (Counts, Buffington, Chang-Rios, Rasmussen, & Preacher, 2010) where the Family Functioning, Social Support, and Nurturing and Attachment scales had Cronbach's alphas of .94, .86, and .83, respectively. In the CQI sample, the Concrete Support scale was more robust (.76) than in the validation study (.63).

Table 13
Cronbach's Alphas for the PFS

Protective Factors Survey	Alpha	If item Deleted
PFS- Family Functioning & Resilience (5 items) (n=234)	0.91	
PFS- Social Support (3 items) (n=234)	0.85	
PFS- Concrete Support (3 items) (n=234)	0.76	.79 if 11
PFS- Nurturing and Attachment (4 items) (n=179) ([a])	0.78	.83 if 14

[a] Only respondents with a child who was eligible for NPSP services answered these items.

Brief Child Abuse Potential Inventory (BCAP)

The Abuse Risk scale's Cronbach's alpha was .80, which shows strong internal consistency. The BCAP had an Abuse Risk scale Cronbach's alpha of .89 in the original validation study (Ondersma et. al., 2005). Cronbach's alphas for the two validity scales, Lie and Random Responding, were not reported in that study. In the current sample, the Lie scale alpha was lower than desired at .62. The Random Responding scale's alpha was negligible in this sample, which indicates these three items did not perform as a coherent scale. A review of the response frequencies for Random Responding revealed almost no variability in scores; only seven participants scored above zero. Other validation studies have demonstrated low internal consistency for the Random Responding scale and recommended that items and the cutoff for this scale should be re-examined (Leil et al., 2019; Walker & Davies, 2012).

Table 14
Cronbach's Alphas for the BCAP

Brief Child Abuse Potential Inventory (BCAP)	Alpha	If item Deleted
BCAP- Abuse Risk (24 items) (n=201)	0.80	
BCAP- Lie (6 items) (n=204)	0.62	
BCAP- Random Responding (3 items) (n=202)	-0.03	

Adolescent-Adult Parenting Index, 2nd Ed. (AAPI-2)

Three of the five AAPI-2 scales had Cronbach's alphas at or above the .70 threshold in this sample. These included Parental Empathy (.72), Corporal Punishment (.85), and Family Roles (.70). Appropriate Expectations had an alpha of .61. If item 1 was dropped, the Cronbach's alpha rose to .65, which indicates the items in this scale are not robust. The Power and Independence scale had a very low alpha of .35. Analyses recommended dropping item 27 to improve the Cronbach's alpha to .44. However, that is still well below the desired threshold and indicates that the items are not measuring Power and Independence coherently or reliably.

Previous research has noted low alphas in the same two scales and with the Family Roles scale, which just meets the threshold in the CQI sample (Connors, Whiteside-Mansell, Deere, Ledet, & Edwards, 2006; sample size = 309). Connors et. al. (2006) found low alphas for Appropriate Expectations (.64 versus .61 in this sample) and Power and Independence (.50 versus .35 in this sample). Further, the Family Roles scale had lower Cronbach's alpha of .59 in the 2006 publication versus .70 in the CQI sample. Both the Connors et. al. (2006) and a more recent study (Lawson, Alameda-Lawson, & Byrnes, 2017) attempted to replicate the 5-factor structure. Connors and colleagues were able to partially replicate the structure, but, as noted above, three of the five factors were not supported. Lawson and colleagues, using a significantly larger sample (n= 2610), also found fair to poor Cronbach's alphas, which mirrors the 2006 study with poor predictive ability to identify future abuse. Additional model analyses indicated that there was a 2-factor structure, but the predictive ability did not improve. A last set of analyses using latent class analysis, found the measure could be very useful in identifying parents who were *highly unlikely* to maltreat their children. As the authors note, this particular ability is most useful in identifying those who do not need intensive services. While these validation studies offer other ways in which the AAPI-2 may be scored and used, these other options are not how the AAPI-2 is currently promoted or used.

Table 15
Cronbach's Alphas for the AAPI-2

Adult & Adolescent Parenting Inventory, 2nd ed. (AAPI-2)	Alpha	If item Deleted
AAPI- Expectations of Children (9 items) (n=155)	0.61	.65 if 1
AAPI- Parental Empathy Towards Child's needs (10 items)(n=155)	0.72	
AAPI- Corporal Punishment (11 items) (n=155)	0.85	
AAPI- Parent Child Family Roles (6 items) (n=155)	0.70	
AAPI- Children's Power and Independence (4 items) (n=155)	0.35	.44 if 27

Psychometrics Summary

Overall, the 57-item FNS, the PFS, and the BCAP showed good internal consistency in this sample, which indicates these measures can reliably produce similar results under consistent conditions. The internal consistency was not as robust in the 25-item FSS in the CQI sample because only two scales show Cronbach's alphas above .70 and five scales with alphas ranging from .51 to .66. The AAPI-2 had one robust scale Cronbach's alpha (Corporal Punishment = .85), two that met the minimum threshold (Parental Empathy = .72 and Parent Child Family Roles = .70), and two undesirable scale Cronbach's alphas (Appropriate Expectations = .61 and Power and Independence = .35).

Measure Correlations

Correlation patterns were examined between all the measures. Three sets of correlations were run and are shown in the next three tables. Table 16 shows correlations of the FNS with the other three measures. Table 17 shows correlations of the PFS with the BCAP and AAPI-2, and Table 18 shows correlations between BCAP and AAPI-2. Several items correlated in expected ways. Statistically significant correlations are highlighted in green.

Table 16

Correlations of the FNS with the PFS (n=216), BCAP (n=187), AAPI-2 (n=162)

	FNS Stress 2-items	FNS- Relationship Discord	FNS- Higher Support	FNS Violence Approval	FNS Fam of O Violence and Neglect	FNS Higher Self Esteem	FNS Depression	FNS Substance Abuse
PFS Family Functioning	-0.24***	-0.44***	0.25***	-0.23***	-0.22***	0.30***	-0.33***	**-0.12**
PFS Social Support	-0.29***	-0.24***	0.44***	-0.15*	-0.20***	0.41***	-0.48***	-0.01
PFS Concrete Support	-0.14***	-0.21***	0.30***	-0.15*	-0.16*	0.23***	-0.35***	-0.12
PFS Nurturing Attachment	-0.07	-0.07	0.10	-0.19*	**-0.13**	**0.13**	-0.18*	0.01
BCAP Risk	0.33***	0.26***	-0.34***	0.07	0.27***	-0.38***	0.54***	-0.04
BCAP Lie	-0.26***	-0.09	0.15	-0.08	-0.27	**0.21*** **	**-0.17*** **	-0.03
BCAP Random	0.05	0.05	0	-0.08	-0.03	0	0	0.02
AAPI- sten Appropriate Expectations	0.01	-0.04	0.05	-0.04	0.01	0.04	-0.06	0.02
AAPI- sten Empathy	0.07	-0.04	0.14	-0.02	-0.07	0.05	-0.11	-0.05
AAPI- sten Corporal Punishment	0.08	-0.13	-0.03	-0.26***	0.01	0.05	0	-0.10
AAPI- sten Family Roles	.-17*	0	0.07	-0.06	0.05	0.06	-0.03	-0.08
AAPI- sten Power Independence	0.10	-0.02	0.03	-0.07	-0.01	0.05	-0.09	-0.07

Note: The means for each subscale were included; Navy has missing items per subscale. Bolded correlations changed when Navy was removed from the sample.

Correlations for the FNS were analyzed in two ways: using the full sample and removing Navy respondents. The second version removed missing data due to items not in the 25-item FSS employed by the Navy. Five scale relationships, of 98 calculated, were significantly changed from the full sample to the smaller sample. These are bolded in Table 15. The two correlations between the PFS Nurturing and Attachment scale and the FNS Family of Origin Violence and Neglect and the FNS Higher Self-esteem scales reached significance in the smaller sample, *r=-.22, p=.05* and *r=.18, p=.05*, respectively. The correlation between the PFS Family Functioning and Resiliency scale and the FNS Substance Abuse scale also reached significance in the smaller sample with an *r=-.21, p=.01*.

Correlations between the BCAP Lie scale and the FNS Higher Self-esteem and FNS Depression scales became significant in the smaller sample, The Lie x Higher Self-esteem correlation changed from *r=.21, p=.001* to *r=-.34*, *p=.05*. The Lie x Depression correlation changed from *r=-.17, p=.001* to *r=.52, p.05*.

The FNS and PFS were correlated across almost all of their scales, and the relationships were in expected directions. Most of the significant correlations were under .4, which indicates modest relationships from one scale to another. Three scales in the PFS had significant correlations to all FNS subscales except for Substance Abuse in the full sample.

- Higher Family Functioning and Resiliency, Social Support, and Concrete Support were each significantly associated with lower reports of Stress, Relationship Discord, Violence Approval, History of Family Violence and Neglect, and Depression in the full sample. As detailed above, the relationship between higher Family Functioning and Resiliency and Substance Abuse became significant in the smaller sample.
- Higher Nurturing and Attachment was significantly associated with lower Violence Approval and Depression in the full sample. In the smaller sample, it was also significantly correlated to lower Family of Origin Violence and higher Self-Esteem.

The FNS and BCAP were also highly correlated with relationships in the expected directions, and the significant correlations ranged between -.38 and .54. Yet, two non-significant correlations are unexpected, and the Random Responding scale had no significant correlations with the FNS. The BCAP Abuse Risk scale was significantly correlated with all but two FNS scales: Violence Approval and Substance Abuse. The Lie scale did not correlate with Higher Support, Violence Approval, or Substance Abuse.

- Higher Abuse Risk was significantly associated with higher Stress, Relationship Discord, Family of Origin Violence and Neglect, and Depression.
- Higher Abuse Risk was also significantly associated with lower Support and healthy Self-Esteem.
- Higher Abuse Risk was not significantly associated with Violence Approval (*r*=.07, *ns*) or Substance Abuse (*r*=-.03, *ns*).
- Higher scores on the Lie Scale, indicating that the respondent was not answering truthfully, were significantly associated with higher Self-Esteem and lower Stress and Depression.

The FNS and AAPI-2 had just two significant correlations, which were small. The non-significant findings show very low *r* values, -.13 to .08 ns, indicating minimal shared measurement of constructs.

- Higher Family Role Reversal scores were significantly associated with higher Stress.
- Lower approval of Corporal Punishment was significantly associated with lower Violence Approval.

The second set of correlations included the PFS, BCAP, and the AAPI-2 (Table 17). The BCAP Abuse Risk scale was significantly correlated to all four PFS scales. The strongest correlations occurred between the BCAP Abuse Risk scale and the PFS Family Functioning & Resiliency and Social Support scales at *r*=-.45, p. 001 and *r*=-.42, p.001, respectively. All correlations between the BCAP Abuse Risk Scale and PFS scales were inverse, which means that as the Abuse Risk scale score increased (higher risk) the PFS scale scores decreased. This indicates that the protective factors were less present or strong. The BCAP Lie scale was also significantly correlated to the Family Functioning and Resiliency and Social Support scales; although, those correlations were modest to weak.

- Lower Abuse Risk was significantly associated with higher Family Functioning, Social Support, Concrete Support, and Nurturing.
- Higher Lie scale scores were associated with higher Family Functioning and higher Social Support, which indicates social desirability may be operating with respect to those two PFS scales. Review of the range of responses on those items shows that they are skewed toward the higher end (more protective factor presence).

The PFS and AAPI-2 had weak to modest correlations (*r*=.16 to .24, all *p.05*) between three of the PFS scales and four of the AAPI-2 scales, for a total of 6 out of 20 possible correlations. Family Functioning was significantly associated with Empathy, Corporal

Punishment, and Power and Independence. Social Support was significantly correlated with Empathy, and Nurturing and Attachment was correlated to Appropriate Expectations and Corporal Punishment. All AAPI-2 and PFS correlations were positive, which indicates lower risk was associated with higher presence of protective factors.

- Higher Family Functioning was significantly associated with higher Empathy and Power and Independence and less approval of Corporal Punishment.
- Higher Social Support was significantly associated with higher Empathy.
- Higher Nurturing and Attachment was significantly associated with more Appropriate Expectations and less likelihood to use or support Corporal Punishment.

Table 17
Correlations of the PFS with the BCAP (n=187), AAPI-2 (n=162)

	PFS Family Functioning	PFS Social Support	PFS Concrete Support	PFS Nurturing Attachment
BCAP Risk	-0.45***	-0.42***	-0.15*	-0.16*
BCAP Lie	0.27***	0.17*	0.05	0.08
BCAP Random	-0.01	0.04	-0.07	0.05
AAPI- Appropriate Expectations	0.01	0.04	0.12	0.21*
AAPI- Empathy	0.21*	0.18*	0.05	0.09
AAPI- Corporal Punishment	0.18*	0.06	-0.06	0.16*
AAPI- Family Roles	0.13	0.08	0.09	0.06
AAPI- Power Independence	0.24*	0.07	0.06	0.08

The third set of correlations compared the BCAP and the AAPI-2. All of the correlations were in the expected direction. However, only two reached significance and were modest to weak. The BCAP Abuse Risk scale was inversely correlated to the AAPI-2 Appropriate Expectations and Empathy scales.

- Lower Abuse Risk was associated with holding more Appropriate Expectations and higher Empathy.

Table 18
Correlations of the BCAP with the AAPI-2 (n=202)

	BCAP Risk	BCAP Lie	BCAP Random
AAPI- Appropriate Expectations	-0.17*	-0.03	-0.08
AAPI- Empathy	-0.19*	0.02	-0.05
AAPI- Corporal Punishment	-0.16	0.14	-0.03
AAPI- Family Roles	-0.07	-0.07	-0.03
AAPI- Power & Independence	-0.13	-0.04	-0.07

Measure Correlation Summary

The CQI measures were all associated with one another at varying levels. Statistically significant associations were in the expected directions. Overall, there were expected patterns of correlations among the FNS, PFS, and BCAP. In addition, the BCAP Abuse Risk scale did not correlate with the AAPI-2 Corporal Punishment scale. The AAPI-2 Empathy scale correlated with at least one scale in each of the other measures, but no other pattern was found with the AAPI-2.

Two non-significant correlations were unexpected. The BCAP Abuse Risk scale did not correlate with the FNS Violence Approval scale or the AAPI-2 Corporal Punishment scale in this sample. The BCAP is specifically designed to screen for child physical abuse risk. A moderate to strong, positive association was expected with voicing approval to use violence in relationships and higher risk in the use of physical punishment. Thus, the lack of finding significant correlations merits further exploration.

The FNS Violence Approval scale has four items, three of which focus on approval of slapping others and one on spanking children. The AAPI-2 Corporal Punishment scale has several items that also focus on the appropriateness of a range of disciplinary actions. In contrast, the BCAP Abuse Risk scale does not ask direct questions about strategies. Rather, the BCAP items about children come from the CAPI's Rigidity scale, which intends to assess rigidity of thinking and family rules/roles.

Perhaps, the FNS and AAPI-2 scales are measuring one's endorsement of physical punishment actions and strategies, while the BCAP Abuse Risk scale is assessing one's ability to be flexible with rules and roles that set up caregiving environments where physical punishment is more or less likely to be used. Although needed, further examination is beyond the capability of the CQI dataset. Thus, future projects may want to consider adding measures of generalized abuse risk to expand understanding of the patterns of convergent validity for these scales.

The BCAP Lie scale correlated with three FNS scales (Stress, Self-Esteem, and Depression) and two PFS scales (Family Functioning/Resiliency and Social Support). These scales have potential for social desirability to skew responses toward the positive. These correlations may indicate that the Lie scale could be useful in understanding other measures and, thus, providing insight into client responses that show a pattern of minimizing troubles or overstating positives.

Measure Discrimination in Identifying Categories of Need (Risk)

Three of the CQI measures are described as screeners that are specifically designed to identify risk: the FNS/FSS, the BCAP, and the AAPI-2. While the risk assessment from the FNS/FSS drives eligibility decisions for home visitation services, this project offered the opportunity to see how these measures compared to the FNS/FSS in categorizing participants into low- and high-needs groups. This section describes how the measures identified lower risk and higher risk groups at Time 1.

FNS High versus Low Needs Identification

Individuals in the CQI sample (n=243) were categorized as either low needs and high needs at Time 1, using the FNS/FSS sum score cutoffs of 9 and 4, respectively, and/or answering at least one of five automatic qualifying items. Thirty-one percent of the sample qualified as high needs due to an automatic qualifying item response. Using the sample with the full FNS (n = 182), 31% were classified as high needs using the cutoff score of 9 or greater. In the Navy-only sample, the cutoff score of 4 classified 59% as high needs.

Five subscales across the PFS, BCAP, and AAPI-2 had scores that were significantly different between the low needs and high needs groups as identified through the automatic qualifying items (Table 19). Individuals in the high-needs category, due to an automatic qualifier, reported lower family functioning and resilience, lower social support, and higher abuse risk and were less likely to trigger the BCAP Lie scale. They also had more family role confusion.

Table 19
FNS/FSS Scale Discrimination for High-Needs Designation Using Automatic Qualifiers (full sample)

Scale	Not FNS High Need (69%)	FNS High Need- Automatic Qualifier (31%)
PFS Family Functioning & Resilience	5.84 (n=152)	5.38 (n=66)
PFS Social Support	6.18 (n=152)	5.81 (n=66)
BCAP Risk	2.65 (n=130)	4.53 (n=59)
BCAP Lie	3.63 (n=130)	3.05 (n=59)
AAPI- Family Roles (raw score)	27.09 (n=111)	29.32 (n=53)

Using the cutoff methods for the FNS and FSS, both screeners differentiated between low and high needs on BCAP Abuse Risk scores (Table 20). The FNS also found differences on scores for the BCAP Lie scale, which indicates high-needs respondents scored lower than their low-needs counterparts and were less likely to trigger that validity scale. The FNS cutoff method also found group differences for all four PFS scales such that those who were not high needs reported more protective factors.

Table 20.
FNS/FSS Scale Discrimination for High-Needs Designation Using Cutoff Scores

	FNS Sample (Excludes Navy)		Navy Only FSS	
Scale	Not FNS High Need (83%)	FNS High Need- 9 or higher (17%)	Not FSS High Need (41%)	FSS High Need- 4 or higher (59%)
BCAP Risk	2.15 (n=92)	5.66 (n=41)	2.20 (n=26)	4.33 (n=33)
BCAP Lie	3.64 (n=92)	2.76 (n=41)		
PFS Family Functioning & Resilience	5.98 (n=110)	5.24 (n=51)		
PFS Social Support	6.48 (n=110)	5.42 (n=51)		
PFS Concrete Support	5.73 (n=110)	5.10 (n=51)		
PFS Nurturing and Attachment	5.18 (n=95)	6.24 (n=32)		

BCAP Low-Risk versus Medium-/ High-Risk Identification

The BCAP calculates risk with lower and upper cutoffs of 9 and 12. Increased risk is 9 or greater and high risk is 12 or greater. Using these cutoffs, the CQI sample (n=204) was

scored into three risk groups (Table 21). The lower cutoff of 9 was used to create two groups, Low Risk and Medium/High Risk, to assess scale score differences.

Table 21
CQI Risk Defined by BCAP Score

BCAP- Risk (n=204)	%
Low Risk(0-8)	90%
Medium Risk (9-11)	4%
High Risk (12+)	6%

Using these risk categories, eight scores across FNS and PFS scales were significantly different between the Low-Risk and Medium-/High-Risk groups (Table 22). Individuals in the Medium-/High-Risk group had higher stress, relationship discord, family of origin violence and neglect, and depression. They also had lower self-esteem, family functioning and resilience, social support, and nurturing and attachment.

Table 22
BCAP Scale Discrimination Using Lower Cutoff of 9

	Full Sample (n=187)	
Scale	Low Risk- BCAP (90%)	Medium/ High Risk BCAP (10%)
FNS Stress (2 items)	1.92 (n=169)	2.81 (n=18)
FNS Relationship discord	1.52 (n=169)	2.03 (n=20)
FNS Family of Origin Violence & Neglect	1.75 (n=171)	2.23 (n=20)
FNS Higher Self Esteem	3.47 (n=170)	2.92 (n=20)
FNS Depression	1.41 (n=170)	2.20 (n=20)
PFS Family Functioning	5.76 (n=181)	4.60 (n=21)
PFS Social Support	6.13 (n=181)	5.11 (n=21)
PFS Nurturing and attachment	6.52 (n=141)	6.27 (n=12)

Two additional steps are required to evaluate whether the Abuse Risk score is valid: the calculations of the Lie scale and Random Responding scale. If scores on either scale

meet or surpass their respective thresholds, the Abuse Risk scale is judged as invalid and not usable in assessing risk for abuse.

In the BCAP validation study (Ondersma et. al., 2005), the invalid protocol rates were 29.7% and 31.9% in the development and cross-validation samples. In contrast, the CQI invalid protocol rates were as follows: 48.3% at Time 1, 56.6% at Time 2, and 55% at Time 3. At each time point, the valid protocol group mean score for Abuse Risk remained stable, while the invalid protocol group's mean score dropped significantly at Times 2 and 3 (Table 23). At all three time points, the mean Abuse Risk scores were significantly different between groups.

Table 23

Mean Abuse Risk Scores from Time 1 to Time 3 x Lie Scale Cutoff

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max	Results
					Lower Bound	Upper Bound			
T1BCAPRisk									
Triggered Lie = No (Valid)	105	4.36	4.272	0.417	3.54	5.19	0	16	$F(1, 201)=23.77$, $p<.001$
Triggered Lie = Yes (Invalid)	98	1.96	2.436	0.246	1.47	2.45	0	14	
Total	203	3.2	3.701	0.26	2.69	3.71	0	16	
T2BCAPRisk									
Triggered Lie = No (Valid)	46	4.33	4.316	0.636	3.04	5.61	0	14	$F(1, 104)=23.29$, $p<.001$
Triggered Lie = Yes (Invalid)	60	1.33	1.856	0.24	0.85	1.81	0	7	
Total	106	2.63	3.484	0.338	1.96	3.3	0	14	
T3BCAPRisk									
Triggered Lie = No (Valid)	27	4.59	4.61	0.887	2.77	6.42	0	14	$F(1, 58)=18.71$, $p<.001$
Triggered Lie = Yes (Invalid)	33	0.94	1.391	0.242	0.45	1.43	0	5	
Total	60	2.58	3.711	0.479	1.62	3.54	0	14	

Three cross-validation studies have reported strengths and limitations of the Abuse Risk scale and challenges for one or more of the validity scales. Walker & Davies (2012) reported alphas of .816 for the Abuse Risk scale and .515 for the Lie scale within a convenience, community-based sample in the United Kingdom. This study also noted that responses on a single item in the Random Responding scale was responsible for a 94.4% invalid protocol rate when scored as recommended in the 2005 validation study. Removing that one item decreased the invalidity rate of the BCAP Abuse Risk score to 30.7%, which was in line with the four samples described in the 2005 study.

Dawe et.al. (2017) assessed the BCAP in a sample of Australian mothers in opioid substitution therapy. Using a different statistical method than the present analyses (Kuder-Richardson Formula 20 versus Cronbach's alpha[4]), the Abuse Risk scale maintained high internal consistency (KR20 = .90), but the Lie scale was unacceptable (KR20 = .05), and the Random Responding scale had poor internal consistency (KR20 = .36). Using the standard scoring procedure, this sample had an invalid protocol rate of 28.7%.

Leil et. al. (2019) tested the validity of the BCAP in a nationally representative German sample of caregivers of children 0 to 3 years old and found that the factor structure could only be confirmed for mothers not for fathers. Leil et. al. (2019) reported strong internal consistency for the Abuse Risk scale with mothers (KR20 = .79), poor for the Lie scale (KR20 = .50), and unacceptable for the Random Responding scale (KR20 = .11). This study also noted that fathers had higher invalid protocols than mothers (30% versus 25%), yet fathers in the invalid protocol group also had higher risk of child maltreatment.

Time 1 analyses of risk status by invalid BCAP protocols revealed broadly similar patterns of invalid protocol groupings, whether using the risk grouping of the FNS automatic qualifier method or the BCAP classification. Table 24 illustrates the groupings using the BCAP's risk-grouping analyses, which had the most conservative estimate of risk (10%) out of the three risk screeners. Comparatively, Table 25 shows where individuals with an invalid BCAP fell when put into the FNS automatic qualifier method risk groups. Analyses for the AAPI-2 risk grouping and invalid BCAP protocols were not run due to the lack of sensitivity of the AAPI-2.

Table 24
Time 1 BCAP Risk Classification x Protocol Status

BCAP Risk Classification	BCAP Valid Protocol	BCAP Invalid Protocol (Lie)
Low Risk (Score: 0-8) (Abuse Risk M=2.16; Lie M=3.58)	43%	False Negative 47%
Medium/High Risk (Score 9+) (Abuse Risk M=12.19; Lie M=2.24)	9%	False Positive 1%

[4] These analyses produce a mathematically equivalent coefficient. The Kuder-Richardson formula provides a shortcut appropriate for binary variables. Cronbach's alpha must be used with continuous variables, but can also be used with binary variables.

Table 25
Time 1 FNS Needs Status (AQ) x BCAP Protocol Status

FNS Needs Status: Automatic Qualifier Method	BCAP Valid Protocol	BCAP Invalid Protocol (Lie)
NOT High Needs (Abuse Risk M=2.65; Lie M = 3.63)	30%	False Negative 39%
High Needs (Abuse Risk M=4.53; Lie M = 3.05)	22%	False Positive 10%

The analyses for risk status by protocol status show that there were a significant number of respondents who were sorted into the low-risk groups by their scores on either the Time 1 BCAP (90%) or FNS (automatic qualifier method: 69%). Yet, there is a substantial contingent in the low-risk groupings whose BCAP scores are invalid due to scoring 4 or higher on the BCAP Lie scale. There is also a smaller contingent of respondents who scored as high needs/risk who also have invalid BCAP scores due to triggering the Lie scale. Cautious explanations about the groupings follow as substantiated child maltreatment data are required to verify whether the groups are identified correctly. It is also worth noting that the researchers who tested the BCAP in the 2006, 2012, 2017, and 2019 studies have not included Milner's original concepts (described below) in explaining different traits of invalid groups in their publications.

The low-risk groups with a valid BCAP are those *who are likely to be truly low needs,* and they answered the BCAP honestly. Likewise, the high-risk groups with a valid BCAP are those *who are likely to be truly high needs* and who answered the BCAP honestly.

The low-risk groups with an invalid BCAP (False Negative) are those who may not be truly low needs. The invalid BCAP indicates that they did not respond truthfully. Note, in Tables 24 and 25, the means for the Lie scale are *higher* for the low-risk groups overall than the high-risk groups. This is likely due to the invalid BCAP group, which is larger than the valid BCAP group, whether using the FNS or the BCAP risk categories. The CAP Inventory's creator (Milner, 1986) describes respondents in this category as "faking good" by answering items in ways that minimize problems, whether through denial or feeling pressure to answer a certain way. As such, the BCAP Abuse Risk score is not trustworthy. These respondents are likely classified erroneously as low risk.

The high-risk groups with an invalid BCAP (False Positive) also did not respond truthfully on the BCAP. While these individuals make up a small proportion of the overall sample, Milner describes some reasons why this might occur when using the full CAP Inventory; although, the CAP Inventory scales used to identify this group are not well represented in the BCAP. A high-risk person with an invalid protocol could be trying to "fake good" and still score as high risk, or a low-risk person could be "faking bad" to try to access resources. In the case of "faking bad," those individuals are likely classified erroneously as high risk. Individuals who tried to "fake good," yet still scored as high risk, would be classified correctly, but every respondent in this group still has a BCAP Abuse Risk score that is not trustworthy.

AAPI-2 Low, Medium, and High-Risk Identification

The AAPI-2 calculates risk using sten scores. There is no overall risk score, but, rather, five separate scores for each of the five scales. Frequencies of risk categorization at Time 1 are listed in Table 26. Using the sten score range, a higher score is better with 1 to 3 being high risk, 4 to 7 medium risk, and 8 to 10 low risk.

Table 26
CQI Risk as Defined by AAPI-2 Scores

Risk Group	Score A Expectations of Children (n=166)	Score B Empathy (n=169)	Score C Corporal Punishment (n=169)	Score D Family Roles (n=169)	Score E Power and Independence (n=169)
High Risk	10%	17%	18%	15%	12%
Medium Risk	64%	57%	56%	44%	39%
Low Risk	9%	10%	7%	19%	31%

Using the risk scores for the AAPI-2, only one significant group difference emerged. The group at low risk for using corporal punishment scored significantly lower on the FNS Violence Approval scale than their high-risk counterparts . The AAPI-2 categories of risk across the 5 scales did not otherwise identify high-risk individuals in this sample.

Table 27
AAPI-2 Discrimination Using Risk Categories

AAPI-2 Corporal Punishment Scale	Full Sample (n=169)	
	Not High Risk	High Risk
FNS Violence Approval	1.30 (n=140)	1.52 (n=29)

High Needs/Risks Discrimination Summary

Three CQI measures were designed to assign individuals into low- or high-needs/risk groups based on those measures' scores. The methods of assigning risk resulted in diverse groupings. The discriminatory performance of the FNS/FSS and BCAP measures closely reflect the measures' correlations in this sample. Table 28 indicates the percentage of participants who were categorized as high needs/risk by each measure and the number of scales across the other measures where there were significant score differences between the low and high groups.

The FNS/FSS was significantly correlated with all four scales of the PFS, the BCAP Abuse Risk scale, Lie scale, and the AAPI-2 Role Reversal and Corporal Punishment scales. Using the FNS cutoff score of ≥ 9, two BCAP and all PFS scale scores were significantly different between low- and high-needs groups. The FSS cutoff score of ≥ 4 produced significant group differences for BCAP Abuse Risk scores. The automatic qualifier method, found group differences for two PFS scales, two BCAP scales, and one AAPI-2 scale.

Table 28
CQI Measure Discrimination for Low /High-Needs Groups

Measure	High Needs/Risk	# Scales: # Measures
FNS/FSS Qualifer Qs	31%	5:3
FNS Cutoff	17%	5:2
FSS Cutoff	59%	1:1
BCAP	10%	8:2
AAPI-2 - Corporal Punishment	18%	1:1

The BCAP Abuse Risk scale was significantly correlated with six FNS scales, all four PFS scales, and the Appropriate Expectations and Empathy AAPI-2 scales. When using the cutoff score of 9, the BCAP identified eight scales across the FNS and PFS where there were significant differences between the low- and high-risk groups it assigned. The BCAP identified 10% of the CQI sample as high needs.

Across the five AAPI-2 scales, there were significant correlations with two FNS scales, all four PFS scales, and the BCAP Risk scale (see Tables 16-18). When using the calculated risk categories across its five scales, only the Corporal Punishment Scale produced group differences for the FNS Violence Approval scale. The AAPI-2 Corporal Punishment scale identified 18% as high needs.

Longitudinal Analyses

Longitudinal analyses completed the assessment of the CQI measures and were used in order to understand if and how each measure might be sensitive to change over time and if that change was in the expected direction. Data across all three time points were included unless the sample size for a given measure at a time point was too small. All findings need to be interpreted with caution as attrition across the time points affects the power of the analyses. Further, analyses linking measures to substantiated cases of maltreatment were not possible. Less than 10 CQI participants met criteria for child maltreatment using the RTC 581 metric.

FNS

Longitudinal analyses were not appropriate to conduct for the 25-item FSS due to a small matched sample size (n=10). Respondents completed the FNS at Time 1 and Time 3 (n=60 matched cases). At Time 1, 31% were classified as high needs by using the automatic qualifier method. The Time 3 group also had 31% who were high needs using the same method. However, 11% of the Time 3 group were also high needs at Time 1.

Using the cutoff score method of classifying FNS high needs, 43% were in the high needs group at Time 3 compared to 17% in Time 1, and 22% of the respondents were classified as high needs at both time points. Table 29 demonstrates how high- and low-needs groups were distributed from Time 1 to Time 3.

Table 29

FNS High Need Classification from Time 1 to Time 3 (n=60)

FNS Score Method	Not High Needs at T1 or T3	Not High Needs at T1 but Qualify at T3	High Needs at T1 but not T3	High Needs at both T1 and T3
Automatic Qualifier	55%	19%	15%	11%
Cutoff Score	46%	19%	13%	22%

While composition of the high-needs group changed from Time 1 to Time 3, no significant differences were found on any of the nine FNS scales between the two data points. Changes in the high-needs group did indicate, however, that need status can and does change. In the small, matched sample, approximately half of respondents had a stable, low-needs classification across both methods. Using the automatic qualifier method, 11% were classified as high needs at both time points, while 22% were if using the cutoff calculations. Percentage of change in status was similar across the cutoff methods.

Using the Time 1 FNS automatic qualifier method, differences existed at Time 3 between the low- and high-needs groups on two scales: The BCAP Abuse Risk and the PFS Nurturing and Attachment scales. The high-needs group scored significantly higher on the BCAP Abuse Risk scale and lower on the PFS Nurturing and Attachment scale at Time 3, which was approximately 7 months after starting NPSP services.

Table 30

Longitudinal Risk: T1 FNS Automatic Qualifier Risk Group and Risk at Time 3

	T1 FNS Automatic Qualifier Not High Need	T1 FNS Automatic Qualifier High Need
T3 BCAP Risk	2.07 (n=44)	4.27 (n=15)
T3 PFS Nurturing and attachment	6.76 (n=50)	6.38 (n=18)

The Time 1 FNS cutoff score method also produced low- and high-needs differences at Time 3. The BCAP Risk Abuse and all four PFS scale scores were significantly different between the low- and high-needs groups. The high-needs group scored significantly higher on the BCAP Abuse Risk scale and lower on each of the PFS scales (Table 31).

Table 31

Longitudinal Risk: T1 FNS Cutoff Method Risk Group and Risk at Time 3

	Only Full FNS (Excludes Navy)	
	Not T1 FNS High Need	**T1 FNS High Need- 9 or higher**
T3 BCAP Risk	1.79 (n=28)	4.11 (n=19)
T3 PFS Family Functioning	5.99 (n=37)	5.34 (n=24)
T3 PFS Social Support	6.37 (n=37)	5.61 (n=24)
T3 PFS Concrete Support	6.02 (n=37)	4.99 (n=24)
T3 PFS Nurturing & Attachment	6.78 (n=35)	6.50 (n=22)

At Time 3, the Time 3 FNS automatic qualifier method identified significant group differences on five scales. The Time 3 AAPI-2 was not included in these analyses due to low sample size. Seven months after services were initiated, the high-needs group at Time 3 scored significantly higher on the BCAP Abuse Risk scale and lower on all four PFS scales than their low-needs counterparts.

Table 32

Concurrent Risk: T3 FNS Automatic Qualifier Risk Group and Risk at Time 3

	T3 FNS Automatic Qualifier Not High Need	T3 FNS Automatic Qualifier High Need
T3 BCAP Risk	1.80 (n=35)	5.60 (n=15)
T3 PFS Family Functioning	5.93 (n=43)	5.21 (n=20)
T3 PFS Social Support	6.44 (n=43)	4.98 (n=20)
T3 PFS Concrete Support	5.89 (n=43)	5.10 (n=20)
T3 PFS Nurturing & Attachment	6.77 (n=41)	6.47 (n=19)

At Time 3, the FNS, with the cutoff score of 9 or greater, method differentiated between low- and high-needs groups on six scales. The high-needs group scored higher on the BCAP Abuse Risk scale and lower on the BCAP Lie scale than the low-needs group (Table 31). The high-needs group also scored lower on all four PFS scales, which indicates the presence of fewer protective factors than the low-needs group.

Table 33

Concurrent Risk: T3 FNS with Cutoff ≥ 9 and Risk at Time 3

	Only Full FNS (Excludes Navy)	
	Not T3 FNS High Need	**T3 FNS High Need- 9 or higher**
T3 BCAP Risk	1.28 (n=25)	4.70 (n=20)
T3 BCAP Lie Score	4.20 (n=25)	2.60 (n=20)
T3 PFS Family Functioning	5.94 (n=34)	5.43 (n=24)
T3 PFS Social Support	6.40 (n=34)	5.50 (n=24)
T3 PFS Concrete Support	5.95 (n=34)	5.11 (n=24)
T3 PFS Nurturing & Attachment	6.82 (n=33)	6.43 (n=22)

PFS

The PFS was completed at Time 1 and Time 3 and 75 respondents completed the measure at both time points. Thirteen additional clients declined to complete the Time 3 PFS. There were no significant differences between Time 1 to Time 3 for any of the four scales.

BCAP

The BCAP was completed at all three time points. While number of respondents dropped from the first to third time point, 98 clients completed both Time 1 and 2, and 60 of those clients completed Time 3. An additional 21 clients declined to complete Time 3. There were significant differences for the Abuse Risk scale between Time 1 and Time 2 and between Time 1 and Time 3. There were no significant differences between Time 2 and Time 3. The Abuse Risk score dropped from Time 1 to Time 2 and remained flat from Time 2 to Time 3. Table 34 shares the paired samples test results for the BCAP.

Table 34

Paired Samples Test T1 to T2 and T1 to T3 BCAP

					95% Confidence Interval				
		Mean	**SD**	**SEM**	**Lower**	**Upper**	**t**	**df**	**sig. (2-tailed)**
Pair 1	BCAPRisk - T2BCAPRisk	0.643	2.79	0.29	0.085	1.20	2.29	97	.02*
Pair 2	BCAPRisk - T3BCAPRisk	0.087	2.77	0.35	0.10	1.51	2.28	56	.03*

AAPI-2

The AAPI-2 was completed at all three time points, using Form A at Times 1 and 3 and Form B at Time 2. Analyses were conducted for Time 1 to Time 2. Analyses with Time 3 data were not possible due to the small sample size. Longitudinal analyses for Time 1 to Time 2 are presented here and should be interpreted with caution. As described in the Measures section, Forms A and B are not equivalent, which is problematic. Analyses were conducted by employing raw scores of scales. Then, analyses were completed again using the sten scores. The results for both sets of analyses were similar. Raw score analyses are presented in Table 35 as there were significant missing sten score data from one of the Services.

There were significant differences from Time 1 to Time 2, about 4 months into services, for Appropriate Expectations, Empathy, and Corporal Punishment. Raw scores improved from 2 to 4 points across these scales, which indicates the clients held more appropriate expectations of children, had higher empathic understanding of children, and were less likely to endorse physical discipline.

Table 35
Paired Samples Test T1 to T2 AAPI-2

		Mean	SD	SEM	Interval Lower	Interval Upper	t	df	tailed)
Pair 1	T1AAP Raw Score A EXPECTATIONS – T2Raw Score A	-2.136	4.46	0.55	-3.23	-1.04	-3.89	65	.00***
Pair 2	T1AAP Raw Score B EMPATHY – T2Raw Score B	-3.606	4.59	0.57	-4.73	-2.48	-6.39	65	.00***
Pair 3	T1AAP Raw Score C CORPORAL PUNISHMENT – T2Raw Score C	-2.818	7.04	0.87	-4.55	-1.09	-3.25	65	.00***
Pair 4	T1AAP Raw Score D FAMILY ROLES – T2Raw Score D	0.273	4.54	0.56	-0.84	1.39	0.49	65	0.63
Pair 5	T1AAP Raw E POWER-INDEPENDENCE – T2Raw Score E	0.348	2.99	0.37	-0.39	1.08	0.95	65	0.35

Analysis Summary

Several types of analyses were conducted with the four CQI measures to understand administration practices, respective reliability and validity properties, how well the screeners distinguished low-from high-needs/risk clients, and if the measures were able to demonstrate change over time. Recommendations and considerations for continued or future use of these measures are presented in the last section of this report, Discussion and Recommendations.

CQI Measure Administration at Time 1

The Clearinghouse team anticipated differences in practice across the Services for the FNS/FSS screener completion. The screener was one of the four measures in the CQI project, so time lag in administering the set could impact the reliability of one or more measures. Analyses identified gaps in administration of the FNS/FSS, ranging from 1 to 4 months, compared to the remaining three measures.

Measure Performance

FNS

The CQI project presented a unique opportunity to test two versions of the military's standardized intake screener: the 57-item FNS and the 35-item FSS. Using the full CQI sample for both models, the longer FNS had higher internal consistencies than the FSS for four of the five comparable scales. Seven of nine FNS scale alphas exceeded the threshold of .70 when recommended items were dropped. Comparatively, only two of six of the FSS scale alphas exceeded the threshold of α =.70 (excluding stress in both models).

The FNS and FSS had very similar correlation patterns with the other three measures, and the FNS analyses had three more significant correlations with PFS scales. Correlations with the BCAP Lie scale indicate that there may be social desirability bias operating with respect to the Higher Self-esteem and Depression scales. These correlations were present in both models.

The FNS and FSS are identical in the five items used to calculate high needs and, thus, equally identified low- and high-needs groups using the automatic qualifier method (31%).

The two screeners have their own cutoff score methods to calculate high needs. and those calculations lead to very different low- and high-needs groups. Using the full CQI sample, the FNS cutoff score identified 17% as high needs, whereas the FSS cutoff score identified 59% as high needs. The three methods of scoring the intake screener resulted in very different needs configurations. The automatic qualifier is used with one or the other cutoff methods, so there are two different ways a person could be identified as high needs from this measure. However, the finding that scoring methods produce different percentages between the low- and high-needs groups is problematic.

Shorter measures are desirable for many reasons. When a shorter version of a measure closely approximates a longer version, organizations may prefer to use the short version. However, the shorter FSS was not as psychometrically robust and identified fewer risk group differences on other measures than the longer FNS.

PFS

The PFS showed high internal consistency across the four scales, and Cronbach's alphas ranged from α = .78 to .91. It performed similarly the validation study sample by the developers. It correlated in expected ways with the FNS/FSS and the BCAP. Each of the PFS scales correlated in expected ways with one or more of the AAPI-2 scales (6 out of 20 possible correlations) at the p=.05 level and with modest correlations (*r= .16 to .24).* Potential for social desirability bias may exist for the Family Functioning/Resilience and Social Support scales as correlations with the BCAP Lie scale indicate.

Of the four CQI measures, the PFS is the only one that does not categorize respondents into risk/non-risk groups. The PFS is in the public domain and requires no contracts or purchasing of rights to use. Its stated purpose is to provide pretest and posttest comparisons for parent education and child maltreatment prevention programs. In this capacity, the PFS did not show any statistically significant change in protective factors from Time 1 to Time 3.

When used in concert with the FNS/FSS at Times 1 and 3, there were significantly different scores on two to four scales for low-needs and high-needs groups (Tables 19 and 20). Used in concert with the BCAP at Times 1 and 3, there were significantly different scores on three of the PFS scales between the low- and medium/high-risk groups (Table 22). This may be a measure that is more useful when paired with another measure that identifies needs/risk groups.

BCAP

The BCAP's primary scale, Abuse Risk, had high internal consistency with α = .80. The Abuse Risk scale had similar internal consistency to the original validation study (Ondersma et. al., 2005) and at least three subsequent, international studies (Dawe, Taplin, & Mattick, 2017; Leil et. al. 2019; Walker & Davies, 2013). Also consistent with the international studies, the CQI analyses demonstrated relative weaknesses of the two validity scales. In this sample, the Lie scale had a Cronbach's alpha of .62, which is below the .70 threshold. The Random Responding scale had a Cronbach's alpha α = -.03, which indicates those items were not related and did not form a coherent scale. The original validation study did not report alphas for the validity scales.

The two validity scales serve to interpret whether a respondent's Abuse Risk scale score is accurate. Total Abuse scores are judged invalid if individuals endorse four or more items on the Lie scale or score greater than 0 on the Random Responding scale. If the internal consistencies of the validity scales are unacceptable to weak, they could adversely influence the Abuse Risk interpretation, which may result in removing accurately identified individuals from high-risk classification and/or incorrectly identifying a high-risk individual as being low risk.

The CQI analyses demonstrated a high rate of invalid BCAP protocols at all three time points, which was significantly higher than the rates in the original validation study and the three diverse, international samples referenced above.

As a screener, the BCAP provided the most conservative estimate of risk at Time 1 (10%) compared to every FNS/FSS calculation method and to the five AAPI-2 scales. This measure is specific to physical child abuse risk, and it is unknown how well it may or may not assess for other forms of abuse or neglect risks.

The BCAP Abuse Risk scale correlated in expected ways, for the most part, with the FNS, PFS, and AAPI-2. Lie scale correlations with scales on the FNS and PFS indicate that it may be able to detect responses on other measures where there is potential for social desirability bias. This finding is in addition to the Lie scale's original role in understanding if an Abuse Risk score is valid. As noted in the analyses, the Abuse Risk scale did not correlate with either the FNS Violence Approval Scale or the AAPI-2 Corporal Punishment scale. Examination of items indicates that the assumption that these would relate may have been in error. The items for the FNS and AAPI-2 scales are similar to one another, and they correlate. However, the Abuse Risk items do not focus on endorsement of punishment strategies or acceptable circumstances for using such actions. Thus, these scales do not demonstrate convergent validity (i.e., measuring the same construct).

The BCAP was completed at all three time points. and analyses show that it is sensitive to change over time as early as about 4 months into services. Respondents scored significantly lower on the Abuse Risk scale from Time 1 to Time 2, and the lower risk was maintained at Time 3. There were high rates of invalid protocols in the low-risk group at each time point, which indicates there may be a significant number of individuals who are minimizing problems and/or are resistant to change.

The 33-item BCAP is a brief version of the proprietary 160-item CAPI. Costs are associated with purchasing inventories on a per administration basis and for scoring the inventories, whether calculated manually or a program is developed to automate scoring. The BCAP does not exist as a purchasable measure. Current recommendations are to purchase the full CAPI, as designed, and then create the BCAP measure and coding system.

AAPI-2

The fives scales in the AAPI-2 varied significantly in their respective internal consistencies, and the lowest performing scales' Cronbach's alphas did not improve sufficiently by dropping an item: Expectations (from .61 to .65) and Children's Power and Independence (from .35 to .44). Two scales met and just exceeded the .70 threshold (Parent-Child Family Roles and Empathy), while the fifth scale demonstrated high internal consistency (Corporal Punishment at .85). The Cronbach's alphas in this sample closely approximated an external validation study of the measure, which indicates two of the five constructs are not measured reliably, and at least one additional scale just meets the minimum needed to demonstrate that the items form a coherent construct.

Additional psychometric concerns were identified in regard to the measure's two forms. There is no identified peer-reviewed research to show that the two forms are equivalent and measure the same constructs in the same ways from Form A to Form B. The forms have only a few shared items, and Form B has more items that are worded negatively. Further, the published manual states that an item that is common across both forms is calculated in Construct E – Children's Power and Independence on Form A and Construct B – Empathy on Form B.

As a screener, one AAPI-2 scale differentiated risk at Time 1 (18%), Corporal Punishment. No other scales identified risk groups by any other measures' scores.

The AAPI-2 scales that were significantly correlated did so in expected ways with the FNS, PFS, and BCAP. Yet, those correlations were modest across the board and occurred less frequently than anticipated.

Time 1 and Time 2 data were analyzed to assess the measure's sensitivity to change over time. The sample size was too small at Time 3 to be included. Form A was completed at Time 1 and Form B at Time 2. Form A was repeated at Time 3. This means that the Time 1 to Time 2 analysis used Forms A and B, and results need to be interpreted cautiously. Three scales showed change over time from Time 1 to Time 2: Expectations, Parental Empathy, and Corporal Punishment. However, the Form A Empathy scale contains the item that is moved into the Power and Independence scale in Form B.

The AAPI-2 is a proprietary measure that must be purchased on a per administration basis. In addition, the developers require that all purchasers use their proprietary scoring and data storage system. There are several concerns about the online system. The scoring formulas are not available independently, and questions about raw score conversion to sten score have not been answered at the time of writing. Data entered into the developer's system cannot be deleted, and there is significant potential for PII/PHI to be entered accidentally. During the course of CQI data collection, website security issues were discovered, which necessitated the halt of data entry until those problems were resolved. The data entry system mirrors the unusual coding key for the 5-point Likert scale: 1-2-4-5-3, which increases the likelihood of data entry error. Recoding entered data could have been handled on the backend of the system instead of putting the burden on the user.

Home Visitor Feedback

The Clearinghouse team solicited feedback from home visitors about the measures used in CQI project. Feedback opportunities occurred after data collection was completed, in fall 2019. Home visitors were sent an anonymous survey link through the email registered on the CQI website. Seventeen of 30 home visitors (57%), who were still active at their installations at the end of the CQI project, responded in part or whole to the survey. Missing responses numbered from 1 to 6 across items. Questions asked about ease of measure administration and usefulness in program planning; resources needed to add CQI tasks into the daily workload; and the usefulness of the project's materials and support connections.

Measure Feedback

Standardized questions were asked about each measure. These questions focused on ease of use and usefulness of the measure in program planning and working with families. Open text opportunities were provided as a way for home visitors to give additional information that may not have been specifically requested and share impressions from clients.

PFS

The majority of responding home visitors (75%; n=12) reported that the PFS was easy to use with families, and it was useful in identifying needs of families (73%; n=11). The CQI portal offered a printout of clients' scores, and 54% of home visitors reported they used the printout. Thirty-nine percent (n=5) of the home visitors who responded indicated satisfaction with the printout resource, while the remaining home visitors reported neutral feelings about it. Half of the home visitors who responded indicated they were able to assess meaningful changes with the PFS over the course of their work with families. Another 14% of home visitors did not find the PFS helpful in assessing change.

Slightly more than 50% of home visitors indicated the PFS identified unique insights about their families, particularly regarding concrete supports and family of origin patterns. Forty-one percent reported that it did not. The majority of respondents (53%; n=9) answered they would recommend the PFS in the future, particularly if it could be integrated into their client management systems.

Ease of use

> *"My clients really enjoyed getting the feedback from the PFS. The color coding helped clients to identify areas that needed shoring up or areas that they were good in. It is worded in a positive manner and I didn't receive any negative comments while administering the survey."*

Usefulness

> *"I like the PFS to use as another tool to helps families see areas they are doing well in, and also areas they strengthen by receiving more services, resources, education, and support. The clients also expressed it was good information to know, and sometimes surprising to see their protective factors were stronger than they expected."*

"I tried to select information and resources to help strengthen the areas that scored yellow or red. In addition, continued to provide them other supportive resources and information to keep the areas strong that scored Green. I also educated the families about the Protective Factors, and how meeting their needs, and knowing how to remain resilient, is beneficial, and strengthens the family even when faced with change, and challenges of daily life."

"The PFS identified specific topics that could have otherwise taken several sessions to assess and identify."

In general, most respondents, ranging from 50-71%, indicated the individual scales were useful for their work with families. For the Family Functioning and Resiliency scale, 69.2%(n=9) reported that they found it very or somewhat helpful, and 15.4% indicated that it was somewhat unhelpful.

"I think it's always good to be able to help families see how they have grown, and become more resilient during times of change, and when faced with stressors."

For the Social Support scale, 71.4% (n=10) of respondents who answered the question indicated that it was very or somewhat helpful, while 7.1% indicated that it was very unhelpful.

"Helped families to talk about their social connections, and ways they can access community services, events, play groups to develop a stronger social support system."

The Concrete Support scale was reported by 69.2% (n=9) of home visitors who indicated this scale was very or somewhat helpful, and 7.7% found it very unhelpful.

"This was probably the most helpful subscale as it targeted specific areas that allowed us to designate specific resources"

The Nurturing and Attachment scale was reported to by 57.1% (n=8) to be very helpful or somewhat helpful, and 7.1% (n=1) indicated it was very unhelpful.

"Assessing bonding and parent's perception of parenting confidence and satisfaction. The FAN can offer support and education if parent identifies a need here. We can discuss protective factors and social emotional development."

"I think most parents answer this in a positive tone, because they feel that is they (sic) way they should answer the questions. They may look like a bad parent if they respond that sometimes they don't like spending time with their child."

When asked if any specific PFS items were notable as being particularly insightful or challenging, home visitors pointed to the family functioning and resilience section. They

felt this section created opportunities to learn more about family dynamics, and families' responses to concrete support items were often surprising. The feedback survey also asked if the PFS helped keep families engaged in services. Home visitors indicated the PFS was a way for families to find validation on how their family functions and to connect to areas where NPSP could provide support. Common responses mentioned the positive wording of the measure as a positive. However, when the PFS was part of a package of measures, concern was noted that it could also contribute to survey fatigue, which made engagement more challenging.

The last question asked if there were any unexpected effects, information, or discussions with families that occurred because the PFS was used. Home visitors reflected that the PFS prompted discussions about family of origin interference or history that negatively affected the current relationship and parenting. Home visitors indicated that this provided an opening for parents to recognize how historical experiences could impact current relationships and to become open to new skill development. In addition, the section on nurturing highlighted that some parents really did not know what to expect from their very young infants. That, then, opened opportunities to share information about developmental milestones.

BCAP

Home visitors were split on the ease of use of this measure. About 31% (n=4) indicated it was somewhat true that BCAP was easy to use. Another 31% (n=4) reported it was not easy to use, and the remaining 39% (n=5) felt neutral about ease of use of the BCAP. A quarter of responding home visitors found the BCAP useful in identifying needs. The majority were neutral (58%, n=7), and 17% (n=2) reported it was not useful in this capacity.

A quarter of home visitors responded that they thought they were able to assess change, which rose to 31% (n=4) when asked specifically about the Abuse Risk scale. A third of home visitors reported they were not able to assess meaningful change, and about 46% felt neutral about its usefulness. About 50% thought the BCAP offered unique insights about families.

There were positive and negative responses when asked if home visitors found the BCAP information helpful in planning visits.

> *I didn't find this tool helpful, and very few of my client's completed it."*
>
> *"It didn't, I reviewed it but didn't find it helpful in planning visits."*

"Typically I found that the BCAP allowed for specific topics of discussion related to parental expectations and if they had unrealistic expectations"

When asked if any specific questions stood out as being insightful or challenging, challenging issues were reported.

"Most of the clients complained about the negative wording on this tool"

"Most of the questions were too strong, and focused on child abuse, also it was more complicated for me to understand how to interpret, and use the results in home visits."

"'Sometimes I have bad thoughts.' Most clients indicate that most people would answer yes to this and would get frustrated at the wording of the questions as if was looking for their faults."

Unexpected effects or discussions occurred with this measure, including clients feeling like they were being set up in regard to the questions. A few clients disengaged from services after completing the BCAP.

Two-thirds (n=7) of the respondents did not recommend using the BCAP in future practice, whether or not it could be integrated into their client management systems.

"This tool is harder to see trends. I would benefit from more training on this tool and its implementation if we begin to use it in practice."

"I am not sure families would be honest in the first 2 visits to answer honestly. I think they also answered according to what they know the "expected answer" should be."

"I believe use of the BCAP may facilitate conversation with the client to assist in the assessment process."

AAPI-2

The majority of responding home visitors (62%, n=8) found the AAPI-2 easy to use with the remaining group feeling neutral about ease of use. Comments on the unusual scale were common. Almost 70% (n=7) indicated that the AAPI-2 was useful in identifying family needs. About 61% also thought they were able to assess meaningful change across data points with this measure, while the other respondents were neutral (39% (n=4)). Just over half of respondents would recommend using the AAPI-2 in future practice. One common

thread in the neutral and negative responses was that this measure is better suited for intervention programming instead of prevention.

Home visitors thought the different scales had different values in program planning.

For the Expectations of Children Scale, 61.5% (n=8) of home visitors, who answered the question, reported the scale was either very or somewhat helpful; 30.8% indicated that they felt neutral about its utility; 7.7% indicated it was somewhat unhelpful.

> *"Providing opportunity to see parental expectations and whether they lined up with normal growth and development."*

For the Parental Empathy Towards Children's Needs Scale, 53.8% reported that it was either very or somewhat helpful; 38.5% indicated they felt neutrally; 7.7% indicated that it was somewhat unhelpful.

> *"If scored low, as a HV I would touch empathy topics."*

For the Use of Corporal Punishment Scale, 61.5% (n=8) indicated that it was either very or somewhat helpful. Another 30.8% were neutral, and 7.7% indicated it was somewhat unhelpful.

> *"Knowing parents (sic) attitudes is helpful. Interventions can be more effective If you know what the underlying reasons for their parenting choices are."*

For the Parent-Child Family Roles Scale, 41.7% (n=5) indicated this scale was either very or somewhat helpful, and 41.7% (n=5) were neutral regarding this scale's utility. The remaining 16.6% indicated that it was either somewhat or very unhelpful.

For the Children's Power and Independence Scale, 53.8% (n=7) responded that it was either very helpful or somewhat helpful, and 38.5% (n=5) were neutral. Similar to other AAPI-2 scales, 7.7% (n=1) indicated that they found it somewhat unhelpful.

When answering the open-ended questions, the items that focused on discipline were insightful to several home visitors. Unexpected responses included learning that parents liked to see how their partners responded and then talk about their differences and similarities, and several parents were surprised at their own scores - usually in a positive way. A few parents felt defensive when reviewing their scores, and these situations challenged home visitors to find ways to reconnect.

Feedback Summary

Overall, the PFS had the strongest support (83%) from home visitors for potentially integrating it into practice. Strong and more varied feedback was given for both the BCAP and AAPI-2. Home visitors and parents found the BCAP uncomfortable (more negative),

and home visitors had more difficultly applying the results to program planning. Only 33% advocated for using the BCAP in future practice. The AAPI-2 had a majority of positive remarks, but there was a sizable neutral group (n of 4 to 5) for almost all aspects of the measure and when examining whether or not the scales were useful for program planning. Fifty-eight percent recommended the AAPI-2 for future use.

Discussion and Recommendations

The NPSP CQI project was completed in two phases over 8 years[5]. In Phase 1, the NPSP Logic Model was updated, and an evaluation plan was supported by all of the Services to be piloted at selected installations. The program's standardized screener and three additional measures were chosen to test the evaluation plan. The foci of Phase 2 were to assess whether any of the additional measures were appropriate for use in continuous quality improvement practices and to gauge the utility of each measure for program planning from a home visitor's perspective.

To these ends, the measures were evaluated for their psychometric qualities, relatedness to one another, abilities to identify needs/risk, and whether they were sensitive to change over time. At the end of the active data collection period, home visitors were asked to reflect and share their thoughts on the additional measures. The analyses of the measures' strengths and limitations did not always align with the home visitors' feedback, but there was common ground.

FNS/FSS

This project provided a unique opportunity to learn how two versions of the standardized screener were used in practice and how they performed from their internal reliabilities to identifying persons eligible for intensive home visitation services. The 57-item FNS is used by the Army, Air Force, and Marine Corps.

The Marine Corps' pilot site used an FNS form that had three additional questions the Marine Corps found useful in assessing eligibility for intensive services. These questions focused on some likely military life/risk experiences: deployment (yes/no; #of

[5] Phase 1 ran from 2012-2014. Phase 2 was completed in 6 years (2014-2020): Two years were needed for the infrastructure build, and the timeline of data collection originally spanned 19 months at each site and included a rolling start system. Data collection was extended to add more sites and participants. See Figure 1 for details.

deployments), self or partner ever diagnosed with PTSD (yes/no), and self or partner ever diagnosed with TBI (yes/no).

The Navy uses a shorter version of the FNS, the 25-item FSS. With the exceptions of slight wording changes on one item and the use of the 4-point Likert scale for responses on involvement in child maltreatment or spouse abuse, the 25 scored items are the same on both versions. Demographic and background data are split between the FSS and a separate form called the Family Social History.

The 57-item FNS performed better in the CQI project in that it showed stronger internal consistency than the FSS. Thus, FNS was a more reliable measure in assessing a client's needs. Both versions performed equally using the automatic qualifier method of assigning need status. Yet, the cutoff score methods produced very different need status groups from one another and from the automatic qualifier method.

The CQI project used the FNS/FSS as a pre-post assessment. One Service currently administers the FNS at 6-month intervals for the duration of a client's participation. FNS analyses show that need status can and does change over time.

Recommendations

The FNS was created in 1999 and was validated with women only. There are at least two versions used in current practice. Given the additional 20 years of research on risk and protective factors for family violence and child maltreatment and almost 20 years of constant war-fighting, refinement of the FNS screening tool is needed. The Marine Corps' additional items and the Navy's shorter version indicate that changes have already occurred, but these modifications are not available or standardized across the Services. Finally, gender is not currently included in the demographic questions of FNS.

The scoring methods produced very different need status groups. In practice, both the automatic qualifier and cutoff scores are used to determine eligibility, but the variability in group identification is concerning.

Recommendation 1: A comprehensive review of the FNS should be completed with the following aims:

- *Identify updates that are relevant to better assess child maltreatment risks in today's military families.*
- *Refine the FNS scoring methods and eligibility criteria.*

There is considerable variation across the Services in when the FNS/FSS is administered to prospective clients and when clients enter into home visitation services. Ideally, the screener would be completed within a month of engaging in services.

Recommendation 2: Screeners, more than 2 months old, should be reviewed by the client and home visitor to ensure updated responses are recorded in the CMS, which ensures current data are driving services.

Within this sample, the 57-item FNS was stronger psychometrically.

Recommendation 3: Analyses support using the 57-item version, limitations withstanding.

PFS

The PFS is a relatively new pretest-posttest measure available in the public domain through the FRIENDS National Resource Center, which is part of the Children's Bureau. It was designed to provide feedback to agencies for program evaluation and continuous improvement of child maltreatment prevention programming. It was used as intended in the CQI project. Psychometrically, the measure performed well, and it correlated in expected ways with the other measures. However, there were no significant changes in scores from Time 1 to Time 3 to show changes in measurable protective factors. When paired with other measures that identify needs status, scores were found to be different between the low- and high-needs/risk groups, but those differences were not visible with just the PFS data.

Home visitors reviewed the PFS very positively, and the majority stated they found the PFS a useful tool to connect with parents and learn information that was otherwise not asked, and they used the information to plan content for visits. The majority of home visitors indicated they would support adding this measure to current practice if the data entry and output were integrated into their client management systems.

Recommendations

The PFS is in the public domain and requires no contracts or purchasing of rights to use, which significantly reduces costs needed to use the measure and to enter, score, and store data. Over 80% of home visitors who provided feedback supported adding the PFS to future practice if the data entry and summary sheets could be integrated into their client management systems.

Recommendation 4: The Protective Factors Survey (PFS) is not recommended as a pre-posttest or as a standalone measure for program improvement.

- *However, the PFS is useful for home visitors in building rapport and using a strengths-based approach in teaching about protective factors.*
- *A set of assessments is required in which at least one of the other measures identifies risk for child maltreatment.*

**It is worth noting that there is now a PFS-2 currently being tested. The measure has undergone significant changes and could be tested for NPSP use in the future.*

BCAP

The BCAP is an adaptation of the CAPI, which is an established screener for risk of child physical abuse. It is one of two measures that showed significant change over time in the CQI project. Analyses showed reduction in risk for abuse at the 4-month mark and that reduction held steady at 7 months. The Lie scale correlated significantly with FNS and PFS scales at Time 1, which indicates it may provide auxiliary value in interpreting scales that are susceptible to social bias responding. As a screener, it gave the most conservative identification of high-risk clients at Time 1. The BCAP is also the measure that was least liked by home visitors and not recommended for future use even if data entry and output were integrated into client management systems.

This is a measure that has ongoing costs associated with it. It is an adaptation of the proprietary CAPI. If length of a measure is an important factor in whether to adopt it, the BCAP may perform similarly to the full 160-item CAPI. Monetary cost will not differ between the two versions, but resource investment to create the BCAP form and the scoring program need to be considered. Home visitors reported that it was difficult to figure out how to use the information it gave, and it was not user friendly as a tool to engage parents.

Recommendations

Recommendation 5: The BCAP demonstrates change in risk over time and could be a useful assessment tool for NPSP when used in concert with other measures such as the FNS.

- *The validity check procedures can help identify potential clients who might otherwise be misclassified into either low- or high-needs groups. In this sample,*

the Lie scale also demonstrated potential to highlight social desirability responses in other measures (the FNS and PFS).

- *If the BCAP is employed, training and support for home visitors and supervisors will be critical, so they can learn how to best use the measure for program planning, discussions with clients, and program improvement.*

AAPI-2

The AAPI-2 is an established, norm-referenced measure that is designed to show change over time and offers two forms that are alternated. It is integrated into a proprietary parent education curriculum but has been used as an independent measure in program evaluation for more than 20 years (original AAPI and AAPI-2). The CQI analyses reinforced other research that indicated at least two of the AAPI-2 scales may have significant internal validity problems. As a screener, it's five scales are scored independently, and each score is assessed for level of risk along a bell curve (low, medium, high). The Corporal Punishment scale is the only one that identified risk differences at Time 1. Three of its scales did show change over time from Time 1 to Time 2. However, these analyses compared Form A to Form B, and there are significant concerns that the forms are not equivalent. An additional concern is that the norm reference groups, which are used to establish the expected range of responses of a specific population, are now more than 20 years old. The manual states that the AAPI-2 currently uses gender and age as norm reference groups.

Home visitors were generally in favor of using the AAPI-2 in future practice. It was already in use at four sites at the time the CQI pilot was implemented.

This is a proprietary measure that has ongoing costs associated with it. In addition to purchasing the measure, the data are entered into a proprietary data management system. Additional concerns were identified about the way data are stored and the risk for PII/PHI exposure.

Recommendations

Several concerns about the AAPI-2 were identified throughout the course of the CQI project including psychometric reliability, performance, and overall data management. Only one concern was identified prior to Phase 2 implementation - the low alphas for three of the five scales in independent studies.

Recommendation 6: Continued or future use of the Adult-Adolescent Parenting Inventory, 2nd Ed. (AAPI-2) is not recommended for the NPSP program. The sunsetting of AAPI-2 involves problems with psychometric reliability, performance, and overall data management.

Summary

The New Parent Support Program was originally selected to participate in a DoD initiative that focused on building capacity for program evaluation. The CQI project produced a common logic model that would be applicable to all the Services' NPSP home visitation programming. An expanded evaluation plan was developed, implemented, and assessed at six pilot installations that represented each Service. Current and potential assessment tools were tested to see how well they performed for program planning and reporting.

Each Service has developed strategies and tools to implement and manage the home visitation program, and some are used in common, while others are tailored to work with local needs and resources. The recommendations from the CQI project have the potential to affect all Services but not all in the same ways. Resource availability and allocation and workforce capacity will need to be reviewed as the recommendations are considered.

References

Bavolek, S. J., & Keene, R. G. (2010). *AAPI online development handbook, 2nd Ed*. Ashville, NC: Family Development Resources, Inc.

Chaffin, M., & Valle, L. (2003). Dynamic prediction characteristics of the Child Abuse Potential Inventory. *Child Abuse & Neglect,* Volume 27, 463-481.

Clearinghouse for Military Family Readiness at Penn State. (2017). *Provider manual: New Parent Support Program CQI pilot project, 3rd ed*. University Park, PA: Author.

Clearinghouse for Military Family Readiness at Penn State. (2017). *Web registration and navigation guide: New Parent Support Program CQI pilot project, 5th ed.* University Park, PA: Author.

Conners, 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, 43*, 762-772.

Dawe, S., Taplin, S., & Mattick, R. P. (2017). Psychometric investigation of the Brief Child Abuse Potential Inventory in mothers on opioid substitution therapy. *Journal of Family Violence, 32,* 341-348.

FRIENDS National Resource Center for Community Based Child Abuse Prevention. (2011). *The Protective Factors Survey user's manual*. Retrieved from http://friendsnrc.org/jdownloads/attachments/pfs_user_manual_revised_2012.pdf

Kantor, G. K., & Straus, M. A. (1999). *Report on the USAF Family Needs Screener*. Durham, N.H.: University of New Hampshire.

Lawson, M. A., Alameda-Lawson, T., & Byrnes, E. (2017). Analyzing the validity of the Adult-Adolescent Parenting Inventory for low-income populations. *Research on Social Work Practice, 27,* 441-455.

Leil, C., Meinck, F., Steinert, J. I., Kindler, H., Lang, K., & Eickhorst, A. (2019). Is the Brief Child Abuse Potential Inventory (BCAPI) a valid measure of child abuse potential among mothers and fathers of young children in Germany? *Child Abuse & Neglect, 88,* 432-444.

Milner, J. S. (1986). *An interpretive manual for the Child Abuse Potential Inventory*. DeKalb, IL: Psytec.

Milner, J. S. (1986). *The Child Abuse Potential Inventory Manual* (2nd ed.). DeKalb, IL: Psytec.

Ondersma, S. J. (2009). *Brief Child Abuse Inventory information sheet*. Author.

Ondersma, S. J., Chaffin, M. J., Mullins, S. M., & LeBreton, J. M. (2005). A brief form of the Child Abuse Potential Inventory: Development and validation. *Journal of Clinical Child & Adolescent Psychology*, *34*, 301-311.

Scott, K. (2004). Final report: Pilot implementation of the Caring Dads program for abusive and at-risk fathers. *Centre for Research on Violence Against Women and Children.* Retrieved from http://www.learningtoendabuse.ca/our-work/pdfs/Pilot_Implementation_Caring_Dads1.pdf

Scott, K., & Crooks, C. (2007). Preliminary evaluation of an intervention program from maltreating fathers. *Brief Treatment and Crisis Intervention, 7,* 224-238.

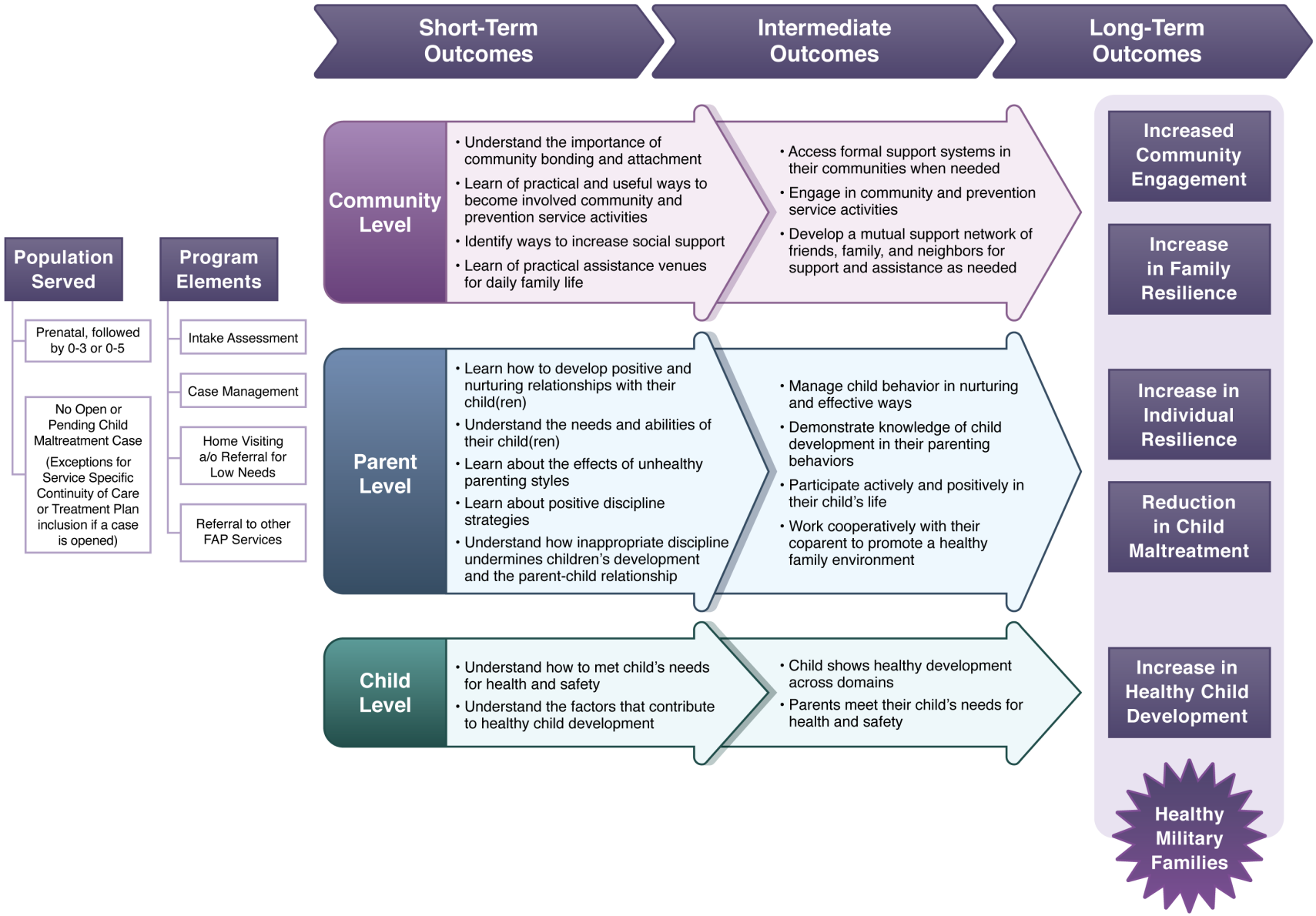
Thomas, D. V., & Looney, S. W. (2004). Effectiveness of a comprehensive psychoeducational intervention with pregnant and parenting adolescents: A pilot study. *Journal of Child and Adolescent Psychiatric Nursing, 17*, 66-77.

Walker, C. A. & Davies, J. (2012). A cross-cultural validation of the Brief Child Abuse Potential Inventory (BCAP). *Journal of Family Violence, 27,* 697-705.

Appendix A

2013 NPSP DoD-Wide Logic Model

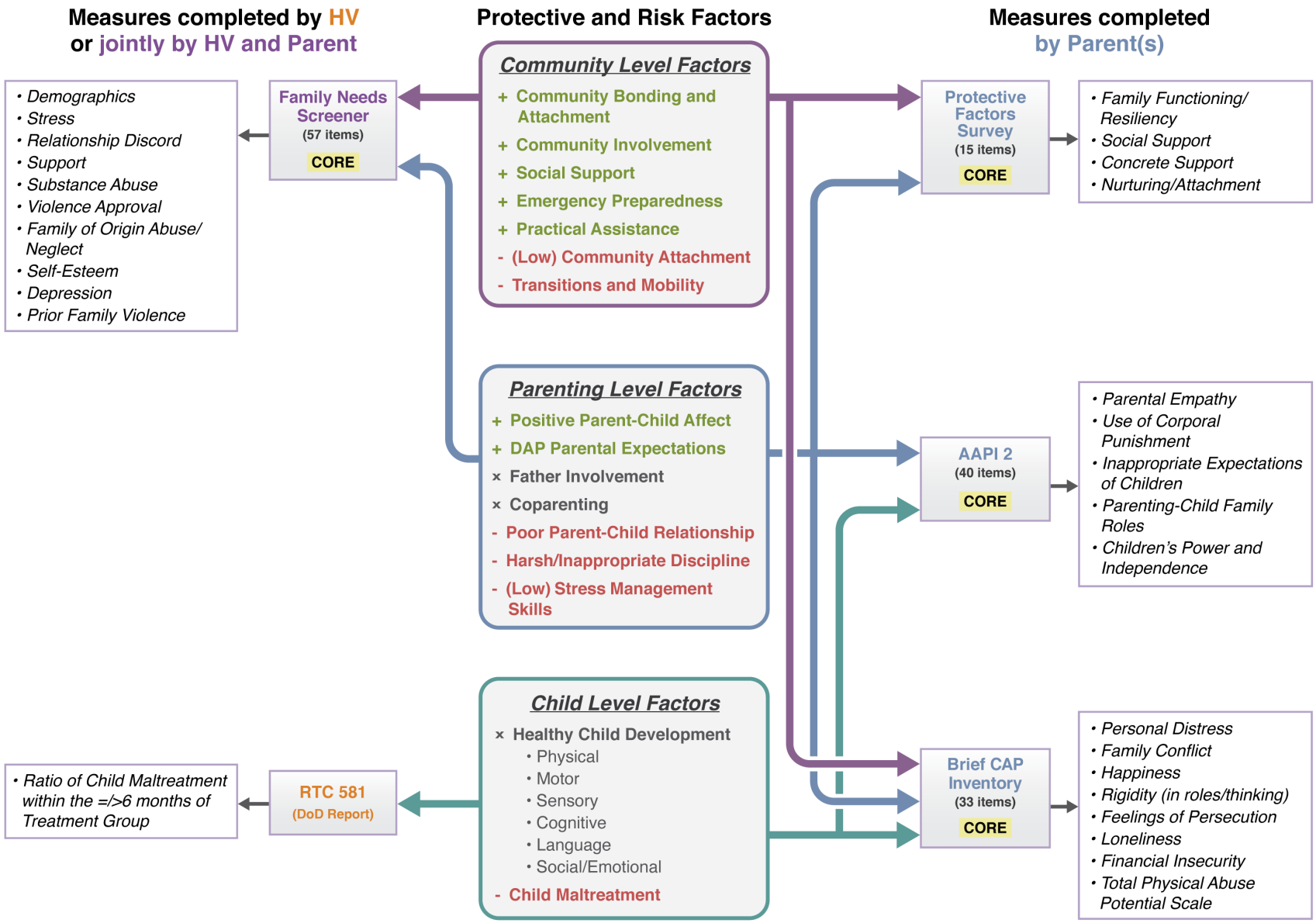
2013 New Parent Support Program Logic Model, DoD-Wide



Appendix B

Tier 1 and Tier 2 Evaluation Models

2013 NPSP Measure Mapping by Protective and Risk Factors: CORE Measures Only Model—REVISED



2013 NPSP Measure Mapping by Protective and Risk Factors: Comprehensive Model—REVISED