Handout on Best Practices for Writing an Evaluation Plan

Example write-ups for evaluation plan components

Component	Sample text
1. Theory of change	When a natural disaster strikes a community, residents can find their homes destroyed and their lives and livelihoods put on hold. Our target population--low-and-moderate income residents--is often the hardest hit by disaster, yet has the fewest resources to return to safe, sanitary, and secure living conditions. Consequently, these residents suffer long-term negative effects, including loss of income, housing instability, and mental health problems from the stress and trauma of the disruption. Our program provides such an intervention by deploying AmeriCorps members to operate a one-stop-shop for residents affected by disaster. AmeriCorps members coordinate volunteers, manage logistics such as material and tool delivery, and oversee construction issues such as permitting and scheduling. We posit that providing a one-stop shop for these services will help low-and-moderate income residents obtain safe and stable housing after a disaster, resulting in greater financial stability, mental and emotional well-being, and a sense of community engagement.
2. Outcomes of interest	The outcome of interest underlying this process evaluation is that disadvantaged individuals are more knowledgeable of the civil legal system (including their eligibility for services, legal options, and/or strategies for navigating the system) as a result of the services provided by AmeriCorps members. This process evaluation will study the preparation and ongoing member support for delivering high quality legal services to underserved individuals using our program's two service delivery models.
	This evaluation will focus on measuring the intended short-term outcomes for members over their service year. Specifically, increases in: English language skills; professional skills; relevant professional & service experience; self-efficacy; and social connections / decreased social isolation
3a. Research questions (process)	This process evaluation will address the following questions: (1) Are the legal services delivery models being implemented with fidelity across sites? Have the recommended enhancements to the service models based on the prior evaluation results been implemented at all sites? If not, why not. (2) Does the training provided by the program and its host sites adequately prepare members to effectively deliver legal services via our program's service delivery models? (3) Do the sites' member management practices (including supervision) adequately support members to effectively deliver legal services via the program's service delivery models?
3b. Research questions (impact)	The primary questions guiding this impact evaluation is: 1. Among parents with prior CPS history, were parents who received program services less likely than parents without such services to experience new allegations of maltreatment?

Component	Sample text
	2. Which factors predict lower likelihood of recurrence, such as prior contact with the child welfare system, demographic characteristics, home visiting program dosage, baseline risk factors as measured by the AAPI, reduction of risk factors as measured by AAPI change scores, or program completion?
3c. Research questions (non-experimental)	The non-experimental evaluation will address the following question: 1.) Have host sites achieved a significant increase in watershed restoration after hosting AmeriCorps members for one service year?
4a. Evaluation design (process/implementation)	The design selected is an internal implementation evaluation using mixed methods. The strength of an implementation evaluation is that it will yield information to help improve the effectiveness of current activities by helping initiate or modify initial activities, provide insight into why certain goals or outcomes are or are not being accomplished, and help program staff make decisions.
4b. Evaluation design (quasi-experimental)	A quasi-experimental design will be used to identify what would happen in the absence of services. A propensity score matching technique will be used to identify a matched comparison group of students who do not receive the tutoring or mentoring services within each school. Students will be matched one-to-one using a nearest neighbor matching algorithm with a .25 standard deviation caliper with exact matching on grade using the following variables: grade, sex, race/ethnicity, English Language Learner status, special education status, free/reduced price meal eligibility and school.
5a. Sampling methods (process & non-experimental)	Due to the small numbers, all supervisors and members will be included in data collection via surveys, interviews and/or pre/post-tests of training. All client satisfaction surveys will be reviewed. No sampling method is needed.
5b. Sampling methods (process)	Sampling for this study is based on the number of AmeriCorps members serving, the number of host site supervisors, and the number of host sites served. All 25 AmeriCorps members submit monthly reports as part of their service requirements and the AmeriCorps Program Director reviews and follows up as needed to make sure the data are complete. The number of host site supervisors in the sample is based on the total number of partners each program year, chosen through a competitive *request for proposal* process. The AmeriCorps Program Manager is charged with following up with all host sites to ensure timely submittal of data. We will use a random numbers table to create a random group of AmeriCorps service members and a group of host sites to serve as our focus group. We will aim to have seven representatives in the focus group which is a common size for a focus group allowing all members to have a chance to comment and share their thoughts and having enough variance in the conversation.
5c. Sampling methods (quasi-experimental)	A power analysis was calculated to ensure that the study is adequately powered to detect significant differences between the intervention and comparison groups. Furthermore, this power analysis enabled the evaluation and program teams to identify whether a sufficient number of children who meet the eligibility criteria would be served by our program. Based on literature from interventions among similar populations, a moderate effect

Component	Sample text
	size (f=0.25) at an alpha of .05 was used to calculate the minimum sample size required to power the study at .80. The power analysis results indicated that a minimum total sample size of 130 total children, or approximately 65 children per group, is required to achieve moderate power. Program management has confirmed that the available pool of children to include in the intervention and comparison groups is sufficient, and meeting the minimum sample size required to properly power the study is feasible. Efforts to include sample sizes larger than the required minimum will ensure that attrition among students does not affect the study's power.
6a. Data (Non-experimental & Process)	*Data Sources* This study will utilize a Pre- and Post-Service Capacity Assessment Survey to measure changes in climate action capacity and goal achievement. Both surveys will be administered online via SurveyMonkey to the intervention and comparison group. - The Pre-Survey will consist of approximately 20 questions, and will include both open- and closed-ended question types. This survey will serve as a baseline measure of local government capacity in a number of climate action domains, including: (1) overall agency capacity, (2) understanding of climate impacts, (3) program and service delivery capacity, (4) stakeholder support, (5) policy awareness, and (6) resource awareness. Under each of these domains, respondents will rate themselves on various items using a 5-point Likert scale. The Pre-Survey will also capture qualitative data about local governments' goals related to climate action, and descriptive characteristics about the respondent and the agency. - The Post-Survey will serve as an endpoint measure of local government capacity and will largely mirror the Pre-Survey. The Post-Survey will also include questions about the agency's achievement of climate action staffing, project, and knowledge sharing goals. - For qualitative data collection, the evaluator will develop a semi-structured interview protocol to guide the beneficiary interviews. Potential interview topics may include: experience in the AmeriCorps program, contributions of the member, outcomes observed at the agency level, perceived impacts of the program at the agency and community level, and program feedback. *Data Collection Procedures* Quantitative data for the study will be collected at two points in time during the 2019/20 service year. The Pre- and Post-Service Capacity Assessment Survey will be administered online via SurveyMonkey to both the intervention and comparison group before the program's service commences (April-July 2019) and again after the service term concludes (June-July 2020). At each administration, agencies will receive up to three reminder emails and two follow-up calls to help ensure a reasonable response rate of 70% or higher on both the pre- and the post-survey. Past administrations of the online pre- and post-surveys among beneficiaries have yielded close to a 100% response rate, which indicates a high response rate can be achieved using a web survey with this

Component	Sample text
	population. The response rate will likely be lower for comparison group agencies. Because the surveys will be administered online, data entry will be performed by the respondent at the time of survey completion. The raw data housed in SurveyMonkey will be exported into a statistical software package for data analysis. Qualitative data for the study will be collected at one point in time at the conclusion of the 2019/20 service year. Based on a list of suggested interviewees developed by the evaluator, the evaluation team will reach out to schedule one-on-one telephone interviews with up to 12 beneficiary agencies. It is expected that the duration of the interviews will be between 30 and 60 minutes each.
6b. Data (Quasi-experimental)	The evaluation will include two data sources: 1. Treatment group data from 9 program sites; and 2. CPS data files from the CPS case records for both treatment and comparison groups. Our program has an extensive database that includes family demographics, dosage of program participation, completion rates, and AAPI assessment data. CPS has data on prior contact, demographics, subsequent CPS contact and the outcome of such contact (substantiated, out of home placement). Our evaluator has an agreement with CPS to receive de-identified data for this evaluation from their records. 1. Extract a data set from our program's database that contains the necessary independent and moderating variables described above for families that were referred by CPS. Send dataset to CPS. 2. CPS isolates a population in its data system within specified time parameters, and within this sample, locate the program clients. CPS will attach the program service data that was submitted for those cases, tag such cases as simply program group cases, and export a de-identified data set to the evaluator which includes CPS data for both the treatment and comparison group regarding CPS priors and subsequent contacts.
7a. Analysis (non-experimental)	For all quantitative data sources, we will perform a visual analysis on the results, building graphs and/or summary tables to gain an initial understanding. This step will also involve disaggregating the results in the following ways: • Members that meet key criteria of our target population vs. members who do not • Members who engaged fully in activities vs. members who did not • Members broken out by their year of service (e.g., first / second / third year) We will leverage an appropriate statistical program (e.g., R or SPSS) to perform analyses and create visualizations. Given the number of expected responses, we anticipate that only tests of statistical significance that include all respondents will be relevant – comparisons between different subgroups of attendees will likely not be statistically significant due to the small sample size. For comparisons of pre and post data, we anticipating using the Wilcoxon Signed Rank test to assess statistical significance (this will likely only be appropriate for the sample as a whole, not subsets of members).
7b. Analysis (process)	For qualitative results obtained from focus groups, we will transcribe participant comments, and summarize results across focus groups thematically with supporting quotes. Given the limited budget of this evaluation and

Component	Sample text
	the number of focus group participants anticipated (10 – 15), we will not be leveraging a specific qualitative analysis software program or developing a comprehensive codebook.
7c. Analysis (quasi-experimental)	Our evaluators will use multiple regression analyses to compare post-measures between treatment and comparison groups to detect any significant differences. Retrospective pre-measures, as well as other variables such as the types of services received, will be included as predictor variables within regression models to control for any pre-existing differences between the groups. The evaluators will assess baseline equivalence between treatment and comparison groups using ANOVA statistical tests. Baseline equivalence will be assessed both for outcome variables and for other variables such as types of services received. We anticipate that given our process for selecting comparison group members any baseline differences will be minimal, but any differences that are detected will be controlled for within our regression analyses.
8. Evaluator qualifications	We plan to engage ABC Consulting as our external evaluator. ABC Consulting is a global research and consulting firm, and is qualified to be our external evaluator because it has experience with: impact evaluation, national service, and disaster services
	We will recruit a lead evaluator and graduate student with the qualifications and skills necessary to collect accurate data, ensure effective data management, and statistically analyze the results of the evaluation study. They will have the program director and program staff at their disposal, with the program director providing an initial program overview and evaluation expectations at the start of the process.
9. Timeline	- Results from a pre-survey will be collected in September and October 2019. We anticipate building this pre-survey into the existing pre-service activities all members are required to complete. Data from this will be analyzed in November 2019, and discussed internally prior to January 2020. - Focus groups will be held near the end of the service year - ideally in April 2020 at the Program's Conference. - Results from the post-survey survey will be collected in May 2020. - Results from this survey will be analyzed in June 2020. - A final report for Year 1 will be drafted during the summer of 2020.
10. Budget	The budget for this evaluation is $x to cover the costs of the evaluator to perform data collection, cleaning and analysis, and a completed evaluation report.