Washington Service Corps - Performance Measurement Training & Technical Assistance (PMT&TA) Program Evaluation

August 2020

Marc Bolan Consulting
www.marcbolanconsulting.com
206 948-0923

EXECUTIVE SUMMARY

The present study investigated the effectiveness of an intervention designed and delivered by the Washington Service Corps to build organizational capacity for program evaluation and performance measurement among AmeriCorps partner sites. A previous external experimental study conducted in 2017 found significant impacts of an earlier version of the intervention on participant self-reported gains in knowledge. With refinements to the WSC intervention and evaluation methods, the present study investigated whether there were greater increases from pre- to post-intervention in a Treatment group vs. a Control group in:

- knowledge about evaluation and performance measurement topics including Evidence of Effectiveness, Program Alignment, Data Collection and Use and High-Quality Data;
- attitudes about and confidence with evaluation and performance measurement topics; and
- capacity to carry out evaluation and performance measurement activities for partner sites in the Treatment group vs. those in the Control group.

The Washington Service Corps (WSC) is a multi-focus intermediary program housed within the Washington State Employment Security Department. As part of the AmeriCorps network, WSC oversees a consortium of approximately 200 partner sites across the state, recruiting, placing, and training between 500 and 600 AmeriCorps Members at these sites each year. WSC operates two major community service programs: 1) WSC AmeriCorps, which places Members at sites to support projects that address needs in focus areas defined by the Corporation for National and Community Service (CNCS); and 2) the Washington Reading Corps, which is dedicated to improving literacy in the state.

Over the past four years WSC developed, tested, and evaluated the Performance Measurement Training & Technical Assistance (PMT&TA) program-- an intervention intended to build the organizational capacity for program evaluation and performance measurement among partner sites. Based on results of the 2017 evaluation study, WSC modified the content and method of delivering this program.

In 2018, WSC contracted with **Marc Bolan Consulting** to conduct a two-year study. The first year (2018-2019) was a pilot test of the modified intervention and refined evaluation tools, and the second year (2019-2020) built upon those findings and used an experimental design to evaluate the implementation and efficacy of the revised program design. The intervention examined in the present study involved a combination of online training webinars, site visits with program staff, customized recommendation reports and follow-up phone technical assistance intended to increase knowledge, change attitudes, and build organizational capacity around evaluation and performance measurement.

In the study, partner sites were randomly assigned to Treatment and Control conditions. The study collected pre- and post- self-report survey data before and after the intervention to measure impacts on participant knowledge of, attitudes about, confidence with, and organizational capacity to implement evaluation and performance measure concepts. Knowledge was measured through multiple choice "quiz" questions and scenarios applying key concepts. The evaluation included qualitative data from participant and staff interviews, as well as "observations" of program site visits, to provide insights on the implementation of the intervention.

Overall, the findings indicated some promising outcomes of the intervention. There were instances of more substantial changes in key outcomes for those in the Treatment group than the Control, and the process evaluation demonstrated that WSC implemented the intervention with fidelity and rigor.

- Treatment group participants demonstrated gains in knowledge of evaluation and performance measurement concepts. Although both the Treatment and Control groups showed increases in overall knowledge over time, the Treatment group showed significantly greater increases in knowledge about a particular concept—Evidence of Effectiveness—than the Control group.
- The Treatment group participants demonstrated greater increases in confidence than Control group participants. There was a significant effect of group assignment (Treatment vs. Control) on changes in the Average Confidence index; confidence increased on many individual items for the Treatment group participants but remained constant or declined for the Control group participants.
- Treatment group participants reported increases in organizational capacity for evaluation and performance measurement, but the Average Capacity Index declined some among Control participants. Treatment group participants reported greater increases over time in capacity in areas such as using and reviewing logic models, creating a detailed plan for data collection, developing a plan to check data quality, and training staff on how to collect and use data. The difference between the two groups may be understated; follow-up interviews with selected Treatment group participants suggested that many organizations intended to carry out evaluation and performance measurement activities, but they were interrupted because of time constraints and the challenges of the Covid-19 pandemic.
- The intervention was implemented consistently and with fidelity and was well-received by participating site staff. Participants reported that WSC was responsive to the needs of their organizations, and that the information provided was timely and useful.
- By using a one-on-one approach with the organizations, WSC was able to tailor the technical assistance to the needs of each organization, helping them better contextualize how the content would support their ongoing evaluation efforts. Moreover, WSC was able to provide sites with specific materials, resources and background research needed to help them advance their evaluation and performance measurement process.

The results of the study suggest the following recommendations for future delivery of this intervention:

- Consider whether the combination of components of the full intervention are necessary if the program is rolled out to a larger number of sites. In particular, there were some questions about the usefulness and efficacy of the online recorded webinar sessions;
- Continue to use a direct one-on-one approach with sites facilitated by experienced WSC staff. WSC might consider how factors such as provision of tools and resources, reduction of barriers to carrying out activities, and communication about WSC expectations can help reinforce positive change over time; and
- Assess the feasibility of tailoring the structure and content of the intervention for different types of organizations or different learning styles; different types of tools or strategies may be more appropriate for particular AmeriCorps focus areas or for newer or more experienced sites or staff.

INTRODUCTION

The Washington Service Corps is a multi-focus intermediary program housed within the Washington State Employment Security Department. The program's mission is to promote and support community-based initiatives that strengthen the capacity of local governments and nonprofit organizations, including faith-based and small community based organizations, to assist individuals, families and communities in need; and to increase the amount of volunteering and civic engagement in Washington state by recruiting, supporting and managing volunteers.

As part of the AmeriCorps network since 1994, WSC oversees a consortium of approximately 200 partner sites across the state of Washington, recruiting, placing, and training between 500 and 600 AmeriCorps Members at these sites each year. WSC operates two major community service programs: 1) WSC AmeriCorps, which places Members at sites throughout the state to support projects that address needs in focus areas defined by the Corporation for National and Community Service (CNCS) as well as other areas of need; and 2) the Washington Reading Corps, which is dedicated to improving literacy in the state.

As part of its requirement to the Corporation for National and Community Service (CNCS), the WSC developed a two-year evaluation plan (2018-2020) to assess the implementation and efficacy of the training intervention intended to build the organizational capacity for program evaluation and performance measurement among partner sites. The WSC's Performance Measurement Training & Technical Assistance (PMT&TA) program was based on materials developed in prior years and combined online training webinars, site visits with program staff, recommendation reports and follow-up phone technical assistance to provide staff with guidance in four topical areas: Program Alignment, Evidence of Effectiveness, Data Collection Use, and High Quality Data.

The original intervention was designed in the Summer and Fall of 2016 and studied across the 2016-2017 program year. To determine the effectiveness of this program, a team of external evaluators from Mission360 Consulting, LLC designed and implemented a randomized controlled trial to measure self-reported changes in staff knowledge, attitudes, and intent to implement strategies learned. The study was conducted across a 12-month period with 32 treatment sites and 32 control sites. A pre-intervention survey was administered in December 2016, WSC delivered the PMT&TA Program during the winter of 2017, and the post-intervention survey was administered in April 2017. Overall, the results showed significant differences in improvements in self-assessed knowledge between the treatment and control groups. The evaluators suggested that a next phase of study could measure actual knowledge changes and shared additional recommendations on ways WSC might refine the program design to address organizational capacity to implement the strategies taught. WSC drew on the findings of the study to modify the structure and components of the intervention.

In 2018, WSC contracted with the evaluators who conducted the present study, **Marc Bolan Consulting**[1], an independent evaluation and research firm based in Seattle, to oversee and manage the evaluation of

[1] Bee's Knees Consulting LLC, an independent research and evaluation firm based in Somerville, MA, provided support in the evaluation design, data collection, and report preparation. Patricia Keenan provided additional data management and data collection support on the project.

the updated training intervention. In the first year (2018-2019), we worked in collaboration with WSC to develop an evaluation plan that would pilot test the intervention using different implementation approaches. In the current year (2019-2020) we drew from the first-year evaluation of pilot sites to develop a plan to assess the efficacy of the intervention using an experimental design comparison group methodology.

Last year's report, titled "Washington Service Corps – Performance Measurement Training and Technical Assistance (PMT&TA) Program Evaluation – Year 1," summarizes the findings from the initial evaluation of the pilot sites. The current report summarizes the data collected and insights gleaned from the second year of program implementation with a particular focus on the impact of the training on a set of outcomes for program staff and organizations. The primary intent of this report is to provide the WSC with the data and insights needed to make decisions about continued use of this intervention. This includes information about effective approaches for connecting with and providing training to the local program staff as well as how to best collect information to assess the implementation and efficacy of the intervention. More importantly, it helps the WSC better understand how the intervention may be effective in building the local capacity of the partner sites.

METHODOLOGY

Description of Intervention and Logic Model

Over the past two years, the Marc Bolan Consulting research staff worked with program and management staff from the WSC in the development, design, and implementation of the evaluation activities. This included the development of a logic model, formulation of a sampling and evaluation plan, development of data collection tools and methods, and data collection. The first-year evaluation was considered a pilot-test, and the focus of the evaluation was on collecting qualitative and quantitative data regarding the implementation of the intervention and the shorter-term outcomes of the intervention. The second year built off the initial findings, and we utilized an evaluation that continued to inform program implementation and included an experimental design methodology to assess whether those receiving the intervention experienced better outcomes than a randomized comparison group.

The original logic model is presented in **APPENDIX 1**. The focus of the model is two-fold, describing the inputs and activities carried out related to the actual implementation of the intervention and those related to the evaluation of the intervention. The short- and medium-term outcomes reflect expected changes in 1) participant knowledge about program evaluation and performance measurement; 2) participants' attitudes about the importance of these processes; and 3) organizational capacity to carry out components of these processes. In the long term, WSC's aim is that their host site organizations will have enhanced ability to implement their own programs and evaluate the efficacy of those programs, using data to learn, improve, and build evidence.

In the second year of the project the Training and Technical Assistance intervention involved a combination of recorded webinars, "site visits" conducted via WebEx, recommendation reports, and a follow-up phone technical assistance session. The first study provided WSC with insights about the appropriate means for delivery of the intervention. Using that information, the staff refined an intervention for the sites in the Treatment group that involved:

- Four recorded webinars presented over a couple of weeks period of time on topics related to Program Alignment, Evidence of Effectiveness, Data Collection Use, and High-Quality Data;
- "Site visits" conducted by WebEx reviewing site-specific evaluation and performance measurement materials delivered by the WSC staff about one month after the completion of the recorded webinars;
- Provision of the recommendation report for that site completed after the "site visit;" and
- A phone follow-up technical assistance session checking in with the site about one month after the provision of the recommendation report.

Sample

At the onset of the program year, the WSC staff identified 55 partner sites as possible candidate sites for the evaluation study. The program staff prioritized sites that did not participate in the last intervention and those likely to have less experience with the training topic areas. For each site, we had information

on the primary staff contacts, the WSC/AmeriCorps focus area designation (e.g., Environmental Stewardship, Disaster Services, Education, Economic Opportunity, etc.), and number of allocated WSC members for the current program year.

WSC designed the intervention to be carried out across three cohorts. This helped the WSC staff to organize and space the timing of the intervention to manage the logistics of scheduling and delivering the various components (e.g., conducting site visits, writing recommendation reports). It also helped the WSC staff to more efficiently gather resources by grouping organizations with similar focus areas together, where possible. We created a stratified random sample assigning sites to the Treatment and Control groups and to each of the three cohorts. We stratified based on focus area and the number of allocated WSC members at the site to ensure similar types of organizations in a cohort and an appropriate mix of smaller and larger organizations in each group in the cohort. The WSC staff then identified up to two staff members at each site to participate in the intervention. These individuals were typically site supervisors, program managers and program line staff, who would likely have more day to day involvement in the program's evaluation activities. **TABLE 1** summarizes the composition of each cohort:

TABLE 1: Study Sample by Cohorts

	Treatment Group		**Control Group**		**Focus Areas**
COHORT	# Sites	# Individuals	# Sites	# Individuals	
A	10	20	10	20	Environmental
B	9	16	8	14	Environmental, Healthy Futures
C	9	17	9	17	Education, Others

To learn more about program implementation, we invited 12 representative sites from the sample above to participate in listening sessions and interviews, to provide us with multiple views into a sample of programs' experiences (4 from each cohort). Of the 12 invited, 7 sites consented, 3 refused, and 2 did not respond to our requests. A member of the evaluation team completed 3 listening sessions (2 from cohort A, and 1 from cohort B); due to WSC WebEx recording and scheduling challenges, we were unable to listen to site visits with the additional 4 sites who had consented. A second member of the evaluation team completed interviews with 13 individuals from all 7 consenting sites, representing all 3 cohorts **(see TABLE 2**). Most interviews included 2 participants from each site at the same time, although 1 program scheduled 2 separate, individual interviews.

TABLE 2: Program Implementation Research Sites

Site	**Cohort**	**Listening Sessions** January 2020	**WSC Recommendation Reports/Follow up Materials**	**Participant Interviews** April-May 2020
1	A	N	Y	Y
2	A	Y	Y	Y
3	A	Y	Y	Y
4	B	Y	Y	Y
5	B	N	Y	Y
6	C	N	Y	Y
7	C	N	Y	Y

Measurement Tools

In **TABLE 3** we present a summary of the different data collection tools used in the study. A more detailed discussion of these tools is presented below.

TABLE 3: Research Study Data Collection Tools

Tool/Method	Data Collection Details	Purpose
Baseline/Follow-Up Participant Survey	Completed by all participants before and after participation in the intervention	Provide background data on participants and data to assess changes in knowledge, attitude and organizational capacity outcomes. The follow-up survey includes feedback data on participation in the full intervention
Participant Interview	Open-ended interviews completed by research staff with participants from seven sites after completion of the intervention	Additional qualitative data regarding perceptions of the intervention and changes experienced over the course of implementation of the training and technical assistance.
Staff Interview	Open-ended interview completed by research staff with key WSC staff member implementing the intervention	Data about successes/challenge associated with delivery of the intervention.
Review of Recommendation Reports	Assessment of content and feedback from recommendation reports completed the WSC from a subset of sites participating in the intervention	Data to inform about the implementation of the intervention and how specific recommendations were shared with the site staff.
"Listening Sessions"	The research staff sat in (virtually) for four site visits	Additional insights about the implementation of the site visit component of the intervention.

The primary data collection instrument used in the study was a survey completed by the participants before starting the webinar component of the intervention (i.e., about one month before) and again at the end of the full intervention (i.e., about one month after). The survey instrument included questions to measure changes in knowledge, attitudes, and organizational capacity outcomes. These are consistent with the short- and medium-term outcomes outlined in the logic model. In the Baseline Survey, we gathered information about the participant's role and experience with the organization, with AmeriCorps, and with program evaluation. In the Follow-up survey, for those in the Treatment sites, we gathered additional information about participants' assessments of the usefulness of the intervention components. The participants completed each survey via online SurveyMonkey platform. The tools are presented in **APPENDICES 2 and 3**.

As noted above, we gathered additional information about the implementation of the intervention by collecting additional qualitative data from a subset of sites. Shortly after the completion of the Baseline Survey, we identified some Treatment sites as possible candidates, where their staff were not as experienced in evaluation and performance measures as other sites, and we could possibly see growth in our key outcome measures. We invited them to participate in the following three activities:

1. "Listening Sessions" – the staff from the site consented to allow a member of the research staff to listen in to their "site visit" with the WSC staff to observe the fidelity to the intended agenda, the use of different materials and activities in the session, the identification of next steps for the

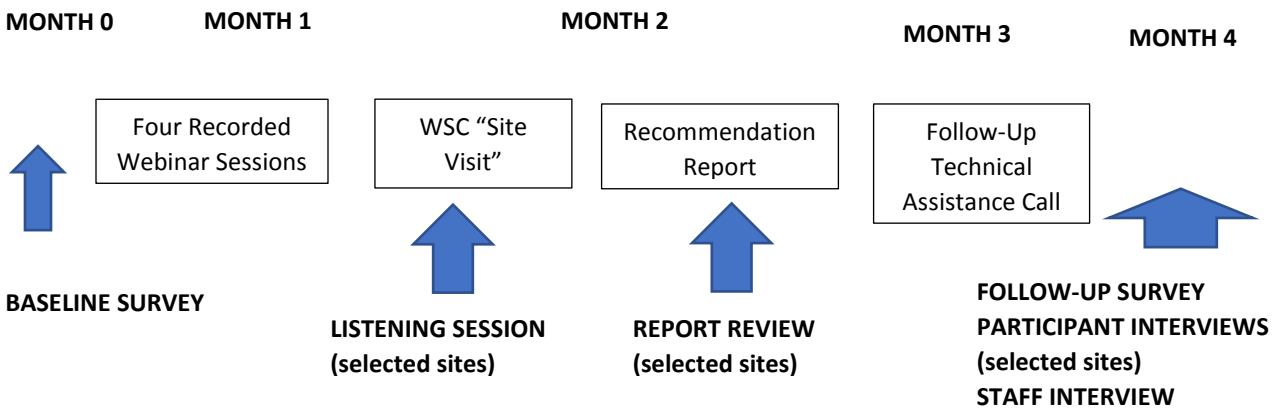
site, the level of engagement among staff participants, and overall successes and challenges with the session. The listening session guide for notes is presented in **APPENDIX 4**;

2. Review of Recommendation Reports – the evaluation team reviewed the WSC reports from these same sites for specific recommendations aligned with specific topic areas, common themes across sites, and information about the provision of additional resources; and
3. Open-ended interviews – after the completion of the full intervention a member of the research team completed an open-ended interview with the staff at these sites. The interview was typically completed with both participants from each site at the same time. The focus of the interview (see **APPENDIX 5**) was on the usefulness of the webinars and other intervention components, on the extent to which participants gained knowledge and changed attitudes about the topic areas, and on the extent to which respondents implemented the skills and recommendations gained through the intervention.

We also completed an interview at the end of the intervention period, in May 2020, with the WSC staff member who implemented the intervention; this staff member carried out the development of the materials and the facilitation of all site visits and follow up calls, and he wrote all of the recommendation reports. Our interview with the WSC staff member (see **APPENDIX 6**) included questions regarding the preparation and implementation of the intervention components with a focus on the site visit and follow-up activities. Additionally, we asked the staff member about some of the successes and challenges with implementation, as well as important issues to consider regarding the sustainability of the intervention and ideas for replication.

FIGURE 1 shows the combination of intervention and data collection components for the evaluation study.

FIGURE 1: Project Intervention and Data Collection Components



Key Research Questions

The intent of the evaluation was to study the impact of the training intervention on short-term participant and site outcomes. As noted, we used an experimental design approach to contrast the changes in outcomes measures for a Treatment and Control group. The primary research questions examined: Does participation in the training intervention result in greater increases pre-post in…

- knowledge about evaluation and performance measurement topics including Evidence of Effectiveness, Program Alignment, Data Collection and Use and High-Quality Data?
- attitudes about and confidence with evaluation and performance measurement topics?
- capacity to carry out evaluation and performance measurement activities?

Quantitative Analytic Methods

The quantitative analyses utilized a combination of descriptive analyses, cross-tabulations, mean comparisons and multivariate approaches to examine change in the key outcomes over time and differences in the patterns of change among those in the Treatment and Control group samples. Our initial descriptive analysis focused on the question of whether there was an attrition bias, particularly in the Treatment group sample, resulting from several individuals not completing the follow-up survey. We further used paired sample t-test analyses to explore changes in key outcomes within the Treatment and Control groups samples and then employed an Analysis of Covariance (ANCOVA) method to examine for differences in the patterns of changes for these measures between the Treatment and Control group samples. We did have 90 total individuals in the paired sample.

Of note, beyond looking at individual survey items we computed a set of indices to provide aggregated data that aligned with our outcomes of interest. See descriptions of these measures in **TABLE 4.**

TABLE 4: Outcome Indices

Index	Definition
Total Knowledge	The survey included seventeen different knowledge questions related to the topics of Evidence of Effectiveness, Program Alignment, Data Collection and Use and High-Quality Data. This index is the number of correct answers on the survey.
Evidence of Effectiveness	The survey included four different knowledge questions related to the topic of Evidence of Effectiveness. This index is the number of correct answers on the survey.
Program Alignment	The survey included four different knowledge questions related to the topic of Program Alignment. This index is the number of correct answers on the survey.
Data Collection and Use	The survey included five different knowledge questions related to the topic of Data Collection and Use. This index is the number of correct answers on the survey.
High Quality Data	The survey included four different knowledge questions related to the topic of High-Quality Data. This index is the number of correct answers on the survey.
Average Attitude	The survey included seven questions where respondents noted their level of agreement with statements about components of the evaluation process on a scale from 1-5 (strongly disagree to strongly agree). The index is the average score (out of 5) across these items with higher scores reflecting more "favorable" attitudes.
Average Confidence	The survey included eleven questions where respondents noted their level of confidence about different components of the evaluation process on a scale from 1-4 (cannot do well to can do very well). The index is the average score (out of 4) across these items with higher scores reflecting more confidence about the ability to use these different components of the process.
Average Capacity	The survey included nine questions where the reported on the extent to which their organization/program has carried out different components of the evaluation process on a scale from 1-4 (No have not done to Yes have done very well). The index is the average score (out of 4) across these items with higher scores reflecting greater demonstrated capacity to implement these activities within the organization

KEY FINDINGS

- **Participants in both the Treatment and Control groups demonstrated gains in knowledge of evaluation and performance measurement concepts over the study period.** There were some variations in this pattern; in particular, we found that Treatment group participants demonstrated more substantial gains in their knowledge regarding Evidence of Effectiveness over time compared to the Control group. WSC covered examples of Evidence of Effectiveness, customized to each Treatment program during their site visits and in their follow up emails, and it could be that the combination of the webinar content with tangible examples of evidence increased the Treatment group's knowledge in this area. The increase in overall knowledge observed for the Control group indicates that individuals from these sites are also picking up knowledge, perhaps a by-product of the standard day-to-day technical assistance WSC provides to partner sites.
- **The Treatment group participants demonstrated greater increases in confidence to implement evaluation and performance measurement activities over time compared to the Control group.** We observed a significant effect of group assignment (Treatment vs. Control) on changes in the Average Confidence index; confidence increased on many individual items for the Treatment group participants but remained constant or declined for the Control group participants. We did not see this variation in the patterns of change for other attitude measures over time; for the most part attitude assessment scores were high at baseline and stayed high for each group.
- **Treatment group participants reported increases in organizational capacity for evaluation and performance measurement, but this outcome declined some for Control group participants.** Although we did not see a significant effect of group assignment on the level of change in the Average Capacity Index, Treatment group participants did report greater capacity in areas such as using and reviewing logic models, creating a detailed plan for data collection, developing a plan to check data quality, and training staff on how to collect and use data at the end of the intervention. Of note, it is possible that the difference in patterns of change might be understated at this point. The follow-up interviews with selected Treatment group participants suggested that agencies were intending to carry out some actions in the future; for some, not enough time had passed since completing the intervention, and for others, the Covid-19 pandemic interrupted their ability to carry out day-to-day operations.
- **Individuals from Environmental Stewardship programs tended to report the greatest gains in measures of organizational capacity.** Environmental organizations composed a sizable percent of the overall study sample, and the analysis showed that participants from environmental programs, in both the Treatment and Control groups, were more likely to demonstrate increases in the Average Capacity Index. During our interviews, a few environmental program staff noted that they had past education in science. It could be that some environmental program staff, who already possessed knowledge about some evaluation concepts, gained new ideas or motivation from the intervention to enhance their evaluation capacity. We also observed in the listening sessions and interviews that WSC provided some environmental programs with some tangible ideas on how to enhance their measurement, so this may have been a contributing factor to this finding.

- **The intervention was implemented consistently and with fidelity to the original intention of the training and technical assistance and was well-received by participating site staff.** The process data showed that nearly all Treatment group participants completed the recorded sessions and subsequently participated in the site visits and follow-up. Additionally, on the follow-up survey, Treatment group participants gave high scores for WSC's communication during the intervention and responsiveness to the needs of the organizations, and they felt that the various components of intervention were very useful. Data from the participant interviews, staff interviews, and listening sessions further indicated that WSC implemented the training faithfully and in a manner that worked well for the participating agencies.
- **Our assessment identified some important contributing factors to the success of the implementation of and potential impact of the intervention.** First, the WSC program staff were effective in working with the participants, communicating the elements of the intervention and helping them understand the importance of the process. Second, by using a one-on-one approach with the organizations, WSC was able to tailor the technical assistance to the needs of each organization, helping them better contextualize how the content would support their ongoing evaluation efforts. Third, through the intervention, WSC was able to provide sites with specific materials, resources and background research needed to help them advance their evaluation and performance measurement process.
- **The recorded webinar knowledge sessions received mixed reviews.** The response to the usefulness of these sessions was mixed among Treatment group participants, with many noting that the information was redundant with what they knew, and the presentation was somewhat dry. When commenting on the most useful components of the intervention and/or their takeaways from the intervention, participants tended to discuss other elements of the intervention (e.g., site visits); few drew on the content and experiences of the webinars.

RECOMMENDATIONS

- The results of this study suggest that it would be beneficial to partner sites to sustain some form of training and technical assistance regarding evaluation and performance measurement over time. Organizations gained knowledge, confidence and capacity through such an intervention and drew from the direct assistance provided via one-on-one site visits, recommendation reports and follow-up. WSC could consider the extent to which such an intervention could be used with a larger set of organizations and whether there is a continued need for the recorded sessions component of the training.
- WSC should further explore the factors that have motivated participants to enhance their confidence and organizations to build their capacity. The results of this study suggest that direct one-on-one work facilitated by experienced WSC staff helped push individuals in a positive direction. For this to continue successfully, the staff should consider how factors such as provision of tools and resources, reduction of barriers to carrying out activities or communication about WSC expectations can help organizations sustain positive change over time.
- WSC might spend more time upfront identifying the right individuals from partner organizations to participate in the training and technical assistance. We did have instances of staff who were likely not the right targets for such a training (i.e., those not a role that involves evaluation) and it is unclear how they best benefitted from the intervention. If a future intervention involves more one-on-one and virtual site visits, this opens the possibility of inviting all staff with the organization who would play a role with the evaluation and performance measurement activities.
- As WSC continues to use this intervention it might be possible to tailor the structure and content even more for different types of organizations or different learning styles. The program might find over time that different types of tools or strategies are more appropriate for particular AmeriCorps focus areas, or that staff can focus on different levels of content for older vs. newer partner sites.
- With respect to continued evaluation WSC should consider gathering data beyond participant self-report. One approach might involve an assessment of the quality of materials provided by the organization to WSC before and after participation in the training. It would be possible to develop a method for comparing elements of things such as logic models, evaluation plans and data collection plans completed before and after to evaluate whether the products they are using have improved over time.

OUTCOME EVALUATION – ASSESSMENT OF EFFICACY OF PROGRAM IMPLEMENTATION

In the analysis, we used survey data from a total of 106 Treatment and Control group respondents from 55 different organizations across the three cohorts. There was some attrition during the Follow-Up Survey assessment period; some individuals did not complete the Follow-Up survey, and other individuals originally assigned to the treatment group did not participate in the intervention so were not invited to complete the Follow-Up Survey. In total, 90 individuals completed the Follow-Up survey, and attrition was higher in the Treatment group sample. The breakdown by cohort and study group is presented in **TABLE 5.**

Additionally, we looked for possible variations in characteristics of the full sample and those in the final paired sample for analysis (n=90). In **TABLES 6-9,** we share these contrasts by different measures from the baseline survey. For the most part, the Paired Sample mirrored the Full Sample, though we did observe that:

- the percent identifying as Primary Site Staff was higher in the Paired Sample;
- the average length of time spent as an AmeriCorps supervisor at the site was higher in the Paired Sample; and
- the percent with greater experience with evaluation and performance measures was higher in the Paired Sample.

TABLE 5: Sample by Cohort and Assignment

	CONTROL		TREATMENT		OVERALL	
COHORT	**Baseline**	**Follow-up**	**Baseline**	**Follow-up**	**Baseline**	**Follow-up**
A	20	18	21	18	41	36
B	14	14	16	12	30	26
C	17	16	18	12	35	28
OVERALL	**51**	**48**	**55**	**42**	**106**	**90**

TABLE 6: Role working with AmeriCorps and WSC

	FULL SAMPLE	PAIRED SAMPLE
Primary Site Staff	62.3%	66.7%
Secondary Site	33.0%	33.3%
Quarterly Staff	27.4%	31.1%

TABLE 7: Samples by WSC Focus Area

	FULL SAMPLE	PAIRED SAMPLE
Environmental Stewardship	47.7%	50.0%
Disaster Services	5.7%	5.6%
Youth Opportunity	5.7%	4.4%
Healthy Futures	14.2%	12.2%
Veterans/Military	1.9%	2.2%
Education	17.9%	17.8%
Economic Opportunity	7.5%	7.8%

TABLE 8: Respondent Experience with WSC and AmeriCorps

	FULL SAMPLE	PAIRED SAMPLE
Average Months (site supported by WSC)	66.9	70.6
Average Months (AmeriCorps Supervisor)	47.7	52.3
Average Months (Associated with AmeriCorps)	87.2	90.9
Served as an AmeriCorps member	61.3%	61.1%

TABLE 9: Experience with Program Evaluation/Performance Measurement

	FULL SAMPLE	PAIRED SAMPLE
I have no or very little experience with any of these activities	6.6%	6.7%
I have some experience with program evaluation	18.9%	14.4%
I have some experience with AmeriCorps performance measures and collecting data for these measures	53.8%	57.8%
I have been involved in all these activities for many years and have expertise in carrying them out!	20.8%	21.1%

As noted, there was substantial attrition of participants in the Treatment group sample. In total, 42 of the original 55 participants (76.4 percent) completed the Follow-up Survey. In **TABLE 10** we explored whether there was any bias resulting from this attrition by comparing those in Paired Sample (n=42) with those who only completed the Baseline survey. We found that individuals in the Paired Sample:

- were far more likely to identify as the Primary Site Staff or the Quarterly Staff;
- were less likely to have served as an AmeriCorps member in the past;
- spent longer periods of time at WSC sites and as an AmeriCorps supervisor;
- had far more past experience with evaluation and AmeriCorps performance measures; and
- had higher baseline Knowledge, Confidence and Capacity indices.

TABLE 10: Attrition Analysis – Treatment Group Sample

	In Paired Sample (n=42)	Baseline Only (n=13)
ROLE		
Primary Site Staff	66.7%	30.8%
Secondary Site	33.3%	30.8%
Quarterly Staff	28.6%	7.7%
Served as an AmeriCorps member	54.8%	61.5%
Average Months (site supported by WSC)	69.6	52.2
Average Months (AmeriCorps Supervisor)	56.8	23.2
Average Months (Associated with AmeriCorps)	88.5	73.8
EXPERIENCE WITH EVALUATION		
No or very little experience	9.5%	7.7%
Some experience with program evaluation	11.9%	38.5%
Some experience with AmeriCorps performance measures	47.6%	38.5%
Involved in all these activities	37.0%	15.4%
Total Knowledge Index	8.50	6.77
Average Attitude Index	4.24	4.27
Average Confidence Index	2.90	2.68
Average Capacity Index	2.76	2.47

Participant Knowledge

In this section, we summarize the survey data regarding measures of participant knowledge of evaluation and performance measurement. The comparison looks at changes in overall indices and individual items, and we used the Analysis of Covariance (ANCOVA) to assess the predictive impacts of sample group assignment.

- Respondents in both the Treatment and Control groups showed increases over time in reported knowledge of evaluation and performance measurement concepts (see **FIGURE 2)**. The Baseline average Total Knowledge Index among Treatment group respondents was higher than the commensurate average for the Control group, but both groups showed significant increases in this Index. The Treatment group average increased from 8.50 to 9.26 (out of a maximum of 17 questions) while the average for the Control group increased from 7.60 to 8.25.
- Looking at the sub-indices that align with the four topic areas addressed in the training and technical assistance, Treatment group respondents showed a significant increase in the Evidence of Effectiveness Knowledge Index but less change over time for the other three indices. In the case of the Control group, there were slight increases over time for each of the four sub-indices **(see FIGURE 3).**

FIGURE 2: Total Knowledge Index: Comparison Over Time – Treatment and Control

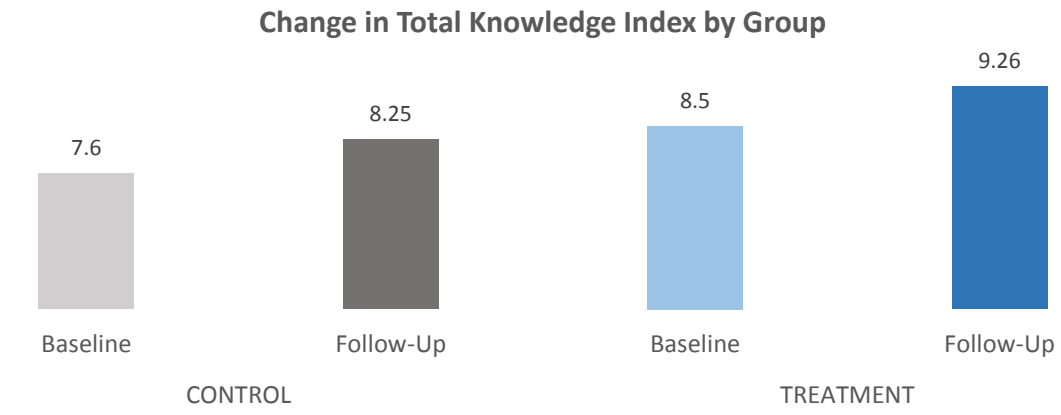
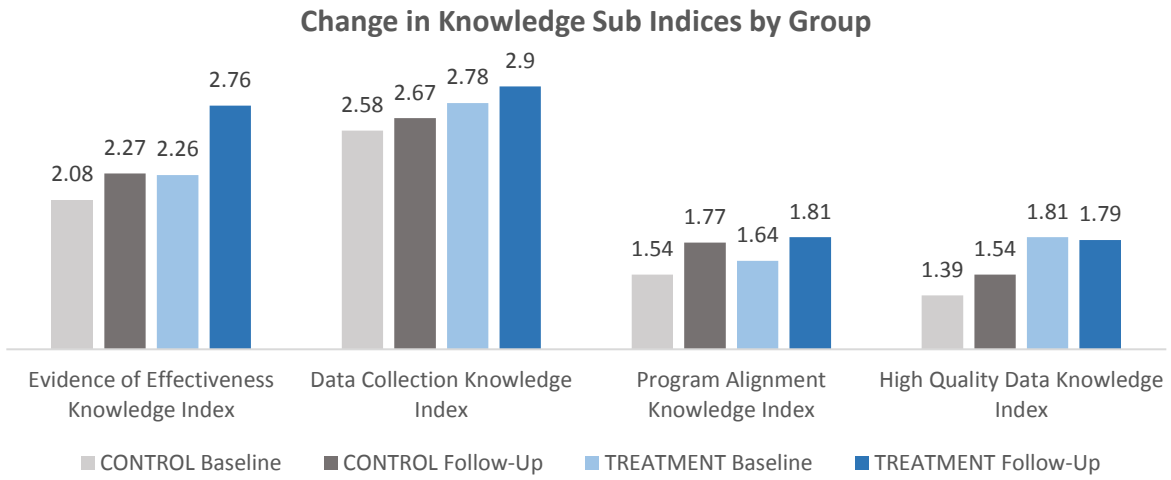


FIGURE 3: Knowledge Sub-Indices: Comparison Over Time – Treatment and Control



As noted, we used the ANCOVA method to evaluate the impact of Treatment vs. Control group assignment on changes in participant knowledge from Baseline to Follow-Up. The previous analysis indicated improvements in the Total Knowledge Index for both samples and some variation in patterns of change for the different sub-indices. Our ANCOVA models considered the impact of group assignment on the Follow-Up knowledge scores, holding constant their Baseline level of knowledge. In **TABLE 10** we present data on the models for the Total Knowledge Index and the five sub-indices. Of note:

- There was no significant effect of Group Assignment on the change in the Total Knowledge Index ($p < .461$). This was consistent with the initial finding showing significant increases in this measure over time for both groups.
- We did see a significant effect of Group Assignment on the change in the Evidence of Effectiveness Knowledge Index ($p < .024$). Although this measure increased for both groups, the change over time was more substantial for the Treatment group participants (i.e., almost .396 units higher change over time when controlling for other factors)
- Group assignment did not appear to be a significant determinant of the change in any of the other knowledge indices.

TABLE 10: ANCOVA Models – Knowledge Outcomes

Assessing Effect of Group Assignment (Control-0, Treatment-1)	**N**	**B**	**SE**	**p-value**
Total Knowledge Index	89	.306	.412	.461
Evidence of Effectiveness Knowledge Index	89	.396	.173	**.024**
Data Collection Knowledge Index	89	.139	.189	.462
Program Alignment Knowledge Index	89	-.049	.192	.800
High Quality Data Knowledge Index	89	.135	.174	.441

*** models control for baseline levels of the outcome and participant's baseline level of experience with evaluation and performance measurement*

The additional analyses presented in this section look at changes over time for each knowledge question for the Treatment and Control group samples (see full tables in **APPENDIX 7**).

Topic Area: Evidence of Effectiveness

- The survey included four questions addressing issues related to Evidence of Effectiveness. As noted earlier in this report, the overall Evidence of Effectiveness Index increased substantially (from 2.26 to 2.76) in the Treatment group and increased in the Control group (from 2.08 to 2.27).
- An analysis of the individual questions provided additional insights on the changes demonstrated over time by both groups: **(see APPENDIX 7: TABLES A1-A4)**
    - With respect to question 1 "evidence of effectiveness of an intervention," we saw that those in the Treatment group were far more likely to report the correct response over time. The percent reporting the correct response increased from 63.4 to 80.5 percent.
    - For question 3 "higher levels of evidence on the continuum," while there was little change in accuracy for the Control group, the percent answering correctly jumped from 54.8 to 83.3 percent among Treatment group participants.
    - In contrast, for question 4 "examples of a study," the percent reporting the correct answer increased more in the Control group than in the Treatment group.

Topic Area: Program Alignment

- The survey included four questions addressing issues related to Program Alignment. Earlier we noted the overall Program of Alignment Index increased (from 1.64 to 1.81) for those in Treatment group and also increased among those in the Control group (from 1.54 to 1.77).
- The analysis of the individual questions provided additional insights on the changes demonstrated over time by both groups **(see APPENDIX 7: TABLES A5-A8)**
    - For Question 1 "parenting workshop logic model," the percent answering correctly increased for both groups but more so among those in the Treatment group. For example, the percent "totally correct" among Treatment group participants increased from 21.4 to 38.1 percent.
    - For question 2 "thorough intervention statement," the increase in knowledge was more evident among the Control group participants; the percent increased from 55.3 to 72.3 percent.
    - Similarly, for question 3 "financial planning workshops" while the percent reporting the correct answers increased for the Control group, this percent decreased for those in the Treatment group.
    - On question 4 "logic model quality," we saw a slight increase in knowledge among those in the Treatment group.

Topic Area: Data Collection and Use

- The survey included five questions addressing issues related to Data Collection and Use. Earlier we noted that the overall Data Collection Knowledge Index increased (from 2.78 to 2.90) for the Treatment group and also increased in the Control group (from 2.58 to 2.67).

- Analysis of the individual questions provided additional insights on the changes demonstrated over time by both groups **(see APPENDIX 7: TABLES A9-A13):**
    - For Question 1 "NOT true about the process of performance measurement," we observed gains in knowledge over time for both Treatment and Control group participants.
    - For Question 2 "appropriate next steps," we saw a somewhat larger decline in knowledge among those in the Treatment group compared to the Control group.
    - For Question 3 "Earth Day event," knowledge increased for those in the Treatment group (i.e., from 76.2 to 83.3 percent correct) but declined for those in the Control group.
    - In Question 4 "change in how someone felt," both groups demonstrated a pretty clear understanding of the correct measure at Baseline and Follow-Up. With the companion question 5, we observed that knowledge in the Control group increased slightly, while this knowledge dropped some for Treatment group participants.

Topic Area: High Quality Data

- The survey included four questions addressing issues related to High Quality Data. Earlier, we noted that the overall High-Quality Data Knowledge Index remained constant (from 1.81 to 1.79) for those in Treatment group yet increased among those in the Control group (from 1.39 to 1.54).
- Analysis of the individual questions provided additional insights on the changes demonstrated over time by both groups **(see APPENDIX 7: TABLES A14-A17)**
    - For question 1 "environmental education workshops," the knowledge appeared to increase over time for the Control group but remained constant or decreased among the Treatment group.
    - For Question 2 "concern with consistency," Treatment group participants showed strong knowledge at both Baseline and Follow-up.
    - For Question 4 "implementation of program services," knowledge increased over time for both groups.

Qualitative Findings: Knowledge

To further understand changes in knowledge over time, we explored data from the open-ended interviews with participants from seven of the Treatment group sites. (Note: a full assessment of the interviews is presented in the process evaluation section of the report). This subset of participants varied in whether or not they felt their knowledge increased as a result of the intervention. Of note:

- About half reported that their knowledge had grown (e.g., one said: "*Knowledge increased in every aspect because the training went so deep into each areas of the entire performance measurement process….[it] made one think of all the possibilities and changes one could make.")* Some examples of knowledge gains mentioned included:
    - Understanding what a supervisor's role looks like and a site's role in collecting data;
    - Learning more about why they collect certain data and how they could make a change (e.g., using a survey and plastic pledge);

- Seeing things from the perspective of WSC and thinking about how they could improve how they obtain useful, quality data;
- Learning more about logic models and how to use them;
- Differentiating outputs and outcomes;
- Gaining clarity in WSC expectations about reporting and ways to improve data accuracy;
- Discovering new evidence/studies that support their program and about the continuum of evidence.

- In contrast, one site noted that their knowledge had not increased much, and two others said it had stayed the same. The reasons they gave for this were generally that they felt the training was general, or some content was redundant/already known by them.

Participant Attitudes

In this section, we summarize quantitative survey data on participant attitudes regarding evaluation and performance measurement. This includes the analysis of agreement items and assessment of confidence items. We examined changes in overall indices and individual items, and we used the Analysis of Covariance (ANCOVA) to assess the predictive impacts of sample group assignment.

- Both groups demonstrated high levels of agreement at both Baseline and Follow-up across many statements regarding evaluation and performance measurement (see **TABLE 11**). The Baseline Average Attitude Index was high for both groups, although it seemed to increase slightly over time for those in the Treatment group (i.e., from 4.31 to 4.44). **(see FIGURE 4)**
- Participants in the Treatment group showed some improvements in attitudes over time, such as increases in levels of agreement with the following items: "We should use evaluation data in the planning and implementation of program services;" "If we had more time and money, I would invest in improving how we collect and use data;" and "We can make necessary decisions about our program services with little or no data *(i.e., higher disagreement over time)*."

TABLE 11: Participant Attitudes Over Time by Treatment and Control Groups

Please assess your level of agreement with the following statements	**CONTROL**		**TREATMENT**	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
Performance measures, or setting goals and measuring progress with data, are a necessary component of our work	4.74	4.70	4.76	4.81
We should use evaluation data in the planning and implementation of program services	4.67	4.72	4.50	4.81
If we had more time and money, I would invest it in improving how we collect and use data	4.15	4.20	3.98	4.17
We can make necessary decisions about our program services with little or no data	1.43	1.52	1.98	1.60
Data about program implementation and outcomes strongly influences our planning for future services	4.30	4.26	4.17	4.17
We should draw from research studies and evidence when developing and planning program services	4.20	4.20	4.43	4.52
The data we gather from participants is only needed for reporting to our funders and interested audiences	1.76	1.83	1.71	1.79

SCALE: 1-strongly disagree TO 5-strongly agree

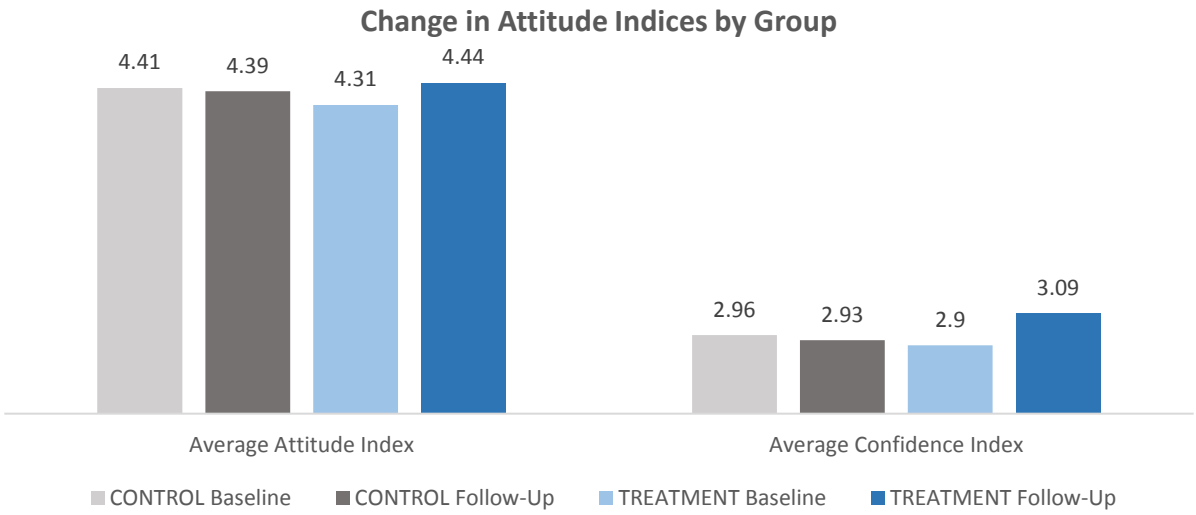
- There were some variations in patterns of change over time when we looked at participants' confidence in carrying out different components of the evaluation process (see **TABLE 12**). In the Treatment group, the Average Confidence Index significantly increased from 2.90 to 3.09 between the Baseline and Follow-Up assessments. In contrast, this index remained constant for those in the Control group (i.e., from 2.96 to 2.93). **(see FIGURE 4)**
- Treatment group participants showed clear increases in self-reported confidence across several items. These included:
    - Creating a logic model
    - Specifying clear outputs, outcomes and performance measures
    - Collecting data in performance measures and outcomes
    - Developing a detailed plan for data collection
    - Developing a method for managing program data
    - Assessing the validity, reliability and consistency of the data

TABLE 12: Participant Confidence Over Time by Treatment and Control Groups

"How would you rate <u>your confidence in your own ability</u> to…."	**CONTROL**		**TREATMENT**	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
Create a program logic model	2.73	2.67	2.64	2.95
Gather information to build evidence for our program	3.00	3.00	3.10	3.10
Specify clear and detailed program outputs, outcomes and performance measures	2.98	2.96	2.90	3.35
Collect data about program performance and outcomes	3.16	3.00	3.17	3.40
Design or identify appropriate data collection tools for the right situation	2.76	2.76	2.76	2.90
Develop a detailed plan for collecting program data	2.89	2.98	2.90	3.17
Develop or identify a formal method to store and/or manage program data	2.93	2.84	2.90	3.21
Analyze and interpret program performance and outcome data	2.96	3.00	3.02	2.95
Assess the validity, reliability and consistency of the data collected	2.96	2.89	2.64	2.95
Identify/use evidence/data to establish the needs of the community	2.84	2.84	2.76	2.83
Use data to inform decision making about the program	3.34	3.30	3.07	3.14

SCALE: 1-cannot do at all TO 4-can do very well

FIGURE 4: Attitude Indices: Comparison Over Time – Treatment and Control



As noted, we used the ANCOVA method to evaluate the impact of Treatment vs. Control group assignment on changes in participant attitudes from Baseline to Follow-Up. Our ANCOVA models considered the impact of the group assignment on the Follow-Up levels of attitudes, holding constant the Baseline attitude level; basically, predicting the impact on the level of change over time. In **TABLE 13** we present data on models for the Average Attitude Index and Average Confidence Index. Of note:

- There was no significant effect of Group Assignment on the change in the Average Attitude Index ($p < .285$). This is consistent with the initial finding showing little change over time in this measure over time for each of the groups.
- There was a significant effect of Group Assignment on the change in the Average Confidence Index ($p < .016$). Our initial analysis clearly showed that this measure only increased for the Treatment group participants, and the ANCOVA model confirms that this change was more substantial for them (i.e., 1.94 units higher change over time when controlling for other factors).

TABLE 13: ANCOVA Models – Attitude Outcomes

Assessing Effect of Group Assignment (Control-0, Treatment-1)	N	B	SE	p-value
Average Attitude Index	87	.095	.088	.285
Average Confidence Index	88	.194	.079	**.016**

*** models control for baseline levels of the outcome and participant's baseline level of experience with evaluation and performance measurement*

Qualitative Findings: Attitudes

We integrated some of the qualitative data from the open-ended interviews with participants from seven of the Treatment group sites into our understanding of changes in attitudes and confidence over time. This subset of participants also varied in the extent to which they felt their confidence had changed. In particular:

- About 4 of the sites noted that they gained confidence because they felt some validation for what they were already doing, because they better understood WSC requirements, or because they understood better how they might make some changes in their metrics and measurement. One participant shared this example, "*In previous [WSC] reports, I would supply the numbers and not be able to explain why we got those numbers and what needs to be done for improvement. I believe I have more confidence going forward*;"
- Two reported that they had no real changes in confidence, and some reasons given included: *"some of it was very 'heady' and not on the ground working stuff"* and "*I have always felt intimidated by statistics, data, and analyses*;" and
- One individual reported feeling "*way less confident*" as a result of the intervention, saying this came from realizing she had been wrong about a lot of information: "*The more we learn, the more we don't know."*

Organizational Capacity

In this section, we summarize the quantitative survey data measuring organizational capacity to implement evaluation and performance measurement activities. The comparison looks at changes in overall indices and individual items, and we used the Analysis of Covariance (ANCOVA) to assess the predictive impacts of sample group assignment.

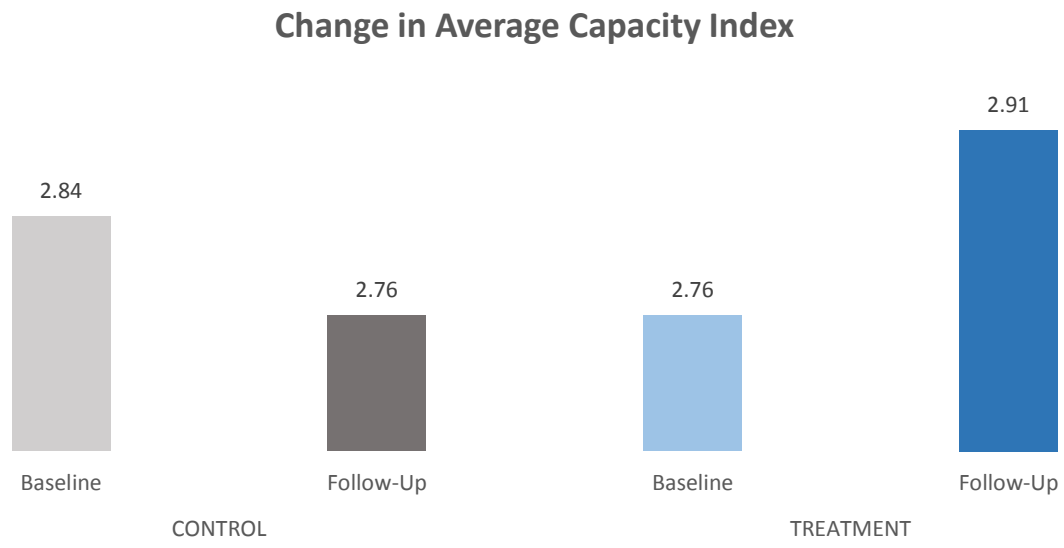
- Respondents reported on the extent to which their program/organization/agency carried out different evaluation and performance related activities (see **TABLE 14**). There is a clear difference in the experience of the Treatment and Control group participants; the Average Capacity Index increased from 2.76 to 2.91 over time among the Treatment group participants while the Index decreased from 2.84 to 2.76 among Control group participants (see **FIGURE 5**).
- Among the Treatment group participants, we observed increases over time for the following behaviors:
    - Using and reviewing the logic model
    - Creating a detailed plan for data collection
    - Developing a plan to check data quality
    - Training staff on how to collect and use data

TABLE 14: Organizational Capacity Over Time by Treatment and Control Groups

To what extent has our program/organization/agency done the following activities:	**CONTROL**		**TREATMENT**	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
Created a complete and detailed logic model that reflects our program	2.82	2.67	2.82	2.91
Used and reviewed our logic model to support decision making and program delivery	2.68	2.52	2.41	2.75
Created a detailed plan for collecting output, outcome and performance measure data	3.10	2.53	2.88	3.10
Developed a plan or strategy for checking the quality of the data that is collected	2.67	2.62	2.43	2.63
Developed or identified a formal method or system to store and/or manage program data	3.07	2.95	3.15	3.24
Use data to inform decision making about the program and/or organization as a whole	3.05	2.87	2.78	2.80
Created a plan to review program data consistently	2.81	2.62	2.58	2.56
Trained program staff on how to collect and use outcome and performance data	2.71	2.71	2.70	2.95
Assigned data collection and evaluation functions to specific people *(e.g., who collects data, who enters it, who reviews it, etc.)*	2.88	2.98	3.10	3.23

SCALE: 1-No We have not done, 2-Yes, we've done of this, but we could do much better, 3- Yes, and we have done a pretty good job with this, 4-Yes, and we have done this very well

FIGURE 5: Average Capacity Index: Comparison Over Time – Treatment and Control



As noted, we used the ANCOVA method to evaluate the impact of Treatment vs. Control group assignment on changes in organizational capacity from Baseline to Follow-Up. Our ANCOVA models considered the impact of the group assignment on the Follow-Up levels of capacity, holding constant the Baseline capacity level. In **TABLE 15** we present data on the model for the Average Capacity Index. In this analysis we found no significant effect of Group Assignment on the change in the Average Capacity Index ($p < .173$). While our initial analysis showed varying patterns of change over time for the

Treatment and Control group participants, this difference was not confirmed in the multivariate analysis when we controlled for baseline levels of the Capacity Index.

TABLE 15: ANCOVA Models – Capacity Outcomes

Assessing Effect of Group Assignment (Control-0, Treatment-1)	N	B	SE	p-value
Average Capacity Index	84	.171	.125	.173

*** models control for baseline levels of the outcome and participant's baseline level of experience with evaluation and performance measurement*

Qualitative Data: Organizational Capacity

We integrated some of the data from the open-ended interviews with participants from seven of the Treatment group sites into our understanding of changes in organizational capacity over time. We found that:

- Sites shared examples of how the intervention resulted in some immediate changes to their evaluation work. In the short-term, about half noted that they had made changes to their reporting and documentation, now that they had greater clarity in how to improve their data quality and how to format it for WSC. For example, one site described how they had implemented an environmental pledge with school education programs and how they shifted to using an online survey (in part because of COVID-19 school closures).
- Participants shared a few reasons why they had not yet implemented some of the learnings. Some noted that short-term plans were interrupted by the COVID-19 pandemic. Other sites reported that they just had not yet had enough time to implement changes they had discussed with WSC.
- Regarding longer term changes, 4 sites mentioned they would be making changes to their survey tools or expanding data collection in new ways (e.g., adapting a survey for broader range of beneficiaries, adding surveys to clean up events, designing tools to better inform decision-making, reviewing their metrics), and 2 described interests in changing their Theory of Change and Logic Model, although the uncertainties due to COVID-19 shutdowns were noted.
- We also asked about the extent to which data the intervention changed how they/their organizations used data.
    - About half felt that they would change/improve how they use data; one site had already begun working with a consultant to build out a schedule for data collection and review, and another felt they would share what they learned with other sites in their organization.
    - Sites who said they would not change how they used data cited bandwidth or pre-existing structures/regulations in their organization (e.g., because they are a federal agency) as reasons. Some named other changes they would make instead (e.g., they would not change how they use data, but they would focus on improving accuracy or think about collecting additional data).

Outcome Comparisons by Experience Level

Additional analyses looked at how the patterns of change for outcome measures may have differed for those who reported varying levels of experience with evaluation and performance measurement at the

Baseline assessment (see **TABLE 16**). We separated the respondent population into Lower Experience (i.e., have No/Little/Some experience with evaluation and performance measurement) and Higher Experience (i.e., some experience with AmeriCorps measures and/or have conducted all activities) groups and then looked at differences for the Treatment and Control samples. While the sample in the Lower Experience groups is smaller, some interesting patterns emerged:

- When looking at changes in Knowledge we found that knowledge typically increased at a higher rate for those with Higher Experience. For example, the Total Knowledge Index increased from 8.63 to 9.49 for Treatment participants with Higher Experience, and similarly we saw a larger increase for the Control groups with Higher Experience.
- We observed increases over time in the Average Attitude Index among Treatment group participants with Higher Experience.
- For the Average Confidence Index, we saw this measure increased for both experience groups in the Treatment sample.
- In looking at the Average Capacity Index we found slightly greater increases over time among the Treatment group participants with Lower Experience than those with Higher Experience.

TABLE 16: Key Outcome Measures by Assignment and Experience

	CONTROL – Lower Exp (n=10)		**CONTROL – Higher Exp (n=38)**		**TREATMENT – Lower Exp (n=9)**		**TREATMENT – Higher Exp (n=33)**	
	Baseline	FU	Baseline	FU	Baseline	FU	Baseline	FU
Total Knowledge Index	6.70	7.00	7.84	8.51	8.00	8.44	8.63	9.49
Average Attitude Index	4.31	4.31	4.43	4.41	4.34	4.32	4.30	4.48
Average Confidence Index	2.60	2.48	3.03	3.03	2.56	2.83	2.99	3.15
Average Capacity Index	2.61	2.52	2.86	2.80	2.73	2.99	2.77	2.88

Comparisons by WSC Focus Area

Additionally, we examined differences in patterns of change for sites/programs with different WSC focus areas (see **TABLE 17**). In the study sample, about 50 percent of the participants worked at an "Environmental Stewardship" organization, so for the purposes of this analysis we divided the Treatment and Control group samples into those worked for "Environmental" organizations vs. "Other"[2] organizations.

- The Total Knowledge Index increased over time for each of the four analytic groups. In the case of the Treatment group, the increase was more substantial over time than for the "Other" organizations.
- Increases in the Average Attitude Index were typically greater for those from "Environmental" organizations, and those from Environmental organizations in the Treatment group showed the greatest increase (i.e., from 4.27 to 4.48).
- There was little variation by Focus area in changes in the Average Confidence Index; this measure increased at similar rates for both groups in the Treatment sample.

[2] Includes Education, Veterans/Military, Healthy Futures, Economic Opportunity, Disaster Services, Youth Opportunity

- There was clear variation in patterns of Capacity/behavior change. The Average Capacity Index only changed for the "Environmental" organizations in both the Treatment and Control samples.

TABLE 17: Key Outcome Measures by Assignment and WSC Focus Area

	CONTROL – Environmental (n=21)		**CONTROL – Other (n=25)**		**TREATMENT – Environmental (n=24)**		**TREATMENT – Other (n=18)**	
	Baseline	FU	Baseline	FU	Baseline	FU	Baseline	FU
Total Knowledge Index	7.52	8.28	7.67	8.22	8.75	9.00	8.17	9.61
Average Attitude Index	4.28	4.37	4.52	4.41	4.27	4.48	4.36	4.40
Average Confidence Index	2.92	2.91	2.98	2.97	2.87	3.08	2.93	3.08
Average Capacity Index	2.81	2.99	2.87	2.56	2.72	2.97	2.83	2.81

PROCESS EVALUATION – ASSESSMENT OF PROGRAM IMPLEMENTATION

Our assessment of the implementation of the intervention drew on sources including the participant Follow-Up survey, participant interviews, staff interview, "listening" sessions, and review of the recommendation reports. The goal was to provide WSC with greater depth of information on how the intervention was carried out, as well as participants' assessments of the strengths and weaknesses of the training and technical assistance.

Overall, looking across our listening sessions, WSC materials and recommendation reports, and input from our WSC staff interview, we observed that WSC delivered the intervention content as intended, following a consistent agenda and set of topics, and sharing materials such as slides, examples of research studies/evidence, and recommendation reports in a consistent manner. Site visits followed a common agenda/structure, reviewing the topics of the webinars and sharing examples relevant or drawn from the individual site/program. The WSC staff member conducting the site visits customized his conversations and materials in response to individual sites' strengths, challenges, and receptivity during their meetings, sensitively and appropriately. Some challenges and variety in implementation across site visits occurred mainly due to technology issues (e.g., internet/WebEx disruptions) and a change in administrative support to the staff member implementing the intervention. We describe some details from each type of data below as well as participant perspectives on the process.

Follow-Up Survey Data

In the follow-up survey, the Treatment group participants answered questions about the training intervention. In **TABLE 18,** we found that participants offered generally positive assessments of the intervention and the WSC staff role in program delivery. Over 92 percent agreed that WSC staff were responsive to "questions and concerns," and over 85 percent agreed they were effective in "communicating content and/or communicating needs and requirements." Of note, over 71 percent also agreed that "it was easy and convenient to participate in the training."

In addition, participants answered open-ended questions about the most and least useful components of the recorded training sessions. The specific responses are presented in **APPENDIX 8**. Participants also commented on suggestions for making these sessions more useful for their organization as well.

These participants completed additional questions about the usefulness of technical assistance components included in the intervention. In general, they found the site visit and recommendation reports more useful (i.e., 80-90 percent reporting somewhat or very useful), and somewhat lower rankings for the phone follow-up visit (i.e., about 62 percent reporting somewhat or very useful (see **TABLE 19**).

The Treatment group participants further commented on any key takeaway actions from the intervention that would help in continued program evaluation efforts as well as suggestions about the additional components of the intervention **(see APPENDIX 9)**. With respect to the takeaways from the intervention, some common themes included:

- Better understanding of the different components of the evaluation process such as development of logic models, evidence of effectiveness and gathering high quality data
- Better understanding of the priorities, needs and expectations of WSC and how to work more collaboratively with the agency on performance measurement issues
- The identification of concrete solutions for specific evaluation issues e.g., identifying the right way to collect data from program participants.

TABLE 18: Assessments of Training Intervention

"Please assess your level of agreement with the following statements"	**Strongly Agree**	**Slightly Agree**	**Slightly Disagree**	**Strongly Disagree**
It was easy and convenient to participate in the training	16.7%	54.8%	21.4%	7.1%
The WSC staff was effective in communicating the content during the recorded training sessions	42.9%	42.9%	14.3%	
The WSC staff was effective in communicating the needs and requirements associated with this training	50.0%	38.1%	9.5%	2.4%
The WSC staff was responsive to our questions and/or concerns about participation in the training	57.1%	35.7%	4.8%	2.4%

TABLE 19: Assessments of Technical Assistance Components

How useful were each of the following WSC Training and Technical Assistance Components	**Not at all**	**A little**	**Somewhat**	**Very**
WebEx Site Visit with the WSC training staff	2.4%	14.6%	39.0%	43.9%
Recommendation Report from WSC after Site Visit	2.4%	7.3%	29.3%	61.0%
Phone/WebEx follow-up program solving session	12.5%	25.0%	42.5%	20.0%

Participant Interviews

As previously discussed, we interviewed staff from 7 sites representing all 3 cohorts. Overall, the participants expressed appreciation for the intervention, and every site shared at least one benefit or takeaway from their participation. They shared a range of feedback and satisfaction with the intervention. Their staff role, educational background, learning style or preferences, and status of WSC funding next year all appeared to shape their impressions of the intervention and perhaps the extent to which it influenced their knowledge, attitudes, behaviors, and future plans for evaluation. Here we share their perspectives on the implementation of the intervention, challenges they face with evaluation, topics they want to learn, and their other recommendations. Their perspectives from these in-depth interviews reinforce the survey findings presented earlier.

<u>Time spent with WSC staff member</u>

- A majority of participants we interviewed described meeting with the WSC staff member, Robert, as the most useful component of the intervention. They felt that the site visit and/or follow up meeting helped them apply the material to the context of their own programs, and they found it helpful to get specific recommendations (e.g., "*All of it was useful, but the big one was the follow up call with Robert," "Talking and working with Robert had an impact,"* "*he made things clearer and helped explain things better…brought it to life;" and* "*The follow up call….was the point in the training where we felt WSC understood us the most….came to a realization of how it all fit together*."

- Only one site found the site visit to be the least useful part of the intervention. They noted that they did not understand why it was called a site visit, that it could have been shorter and more efficient, which they found frustrating. They felt an in-person visit would have allowed WSC to see their program in action and help reduce time they lost explaining their program.

Recommendation reports

- A couple of participants noted that the recommendation reports were particularly useful (e.g., *"Even when they confirmed something I already knew, they were fruitful, productive, and useful*" [and included examples of logic models or other tools they could use]. Another noted that the recommendations confirmed some plans or recommendations they had developed elsewhere, such as using Google forms/electronic surveys and doing a comparison study. They also appreciated evidence from other research studies that WSC shared.

Webinars/Recorded content

- A majority of sites (6 of 7) interviewed said they considered webinars/recorded content to be the least useful element of the intervention. Examples of comments include, "*Sometimes watching the webinar videos, you can kind of lose interest and have to pull yourself back in…it was a lot to take in;" "Not the biggest fan of the webinars—it was a little bit of the 'death by PowerPoint," "this part was just a review of information and not really helpful,"* "*really long and general,"* and "[some voicing was] *monotone."*
- Yet, some participants felt the content was new and expressed appreciation for the webinars (e.g., "*The online training was very helpful, especially in the first one—it was very concrete and increased knowledge about the differences between outputs and outcomes,"* and they named specific webinars that helped them (e.g., Evidence of Effectiveness and Program Alignment). One participant felt the practice sessions and exercises were very useful because then they could relate the webinar content to their own program, and she appreciated the ease/flexibility of being able to watch the webinars at times that worked for them.

Challenges to organizational capacity to implement evaluation/measurement

- The most common challenge named by sites was time/staff capacity (4 of 7 sites noted this). In addition to not having enough time or staff, they described related issues such as turnover and lack of role clarity within their organization.
- Technology was another challenge that multiple sites noted, and the COVID-19 pandemic had underscored it. Although some programs had shifted to online surveys, for example when schools closed, not all of their beneficiaries have access to technology (e.g., internet).
- Other miscellaneous challenges noted by individuals included: keeping a schedule; the unique nature of their program and the range of services, outcomes, or ages they cover; obtaining parent consents for student surveys; collecting demographic data in non-discriminatory ways;

Topics of interest/suggestions for future.

- We asked sites to share any evaluation-related topics they wish could have been covered or that they would like to learn more about or recommendations for the future. The input below represents a range of ideas shared by at least one individual; there was not a consistent theme, but some may spark future innovations or just show the range of their interests. Topics noted included:
    - Outcomes for AmeriCorps members
    - On the ground examples of how to conduct an evaluation study

    - How to make changes to measurement midstream and still obtain useful and accurate data
    - Help in crafting surveys
    - How to integrate evaluation/data collection into programming so it does not feel disruptive to students
    - How to measure behavior change
    - How to collect qualitative data
    - How to apply information on tool development to their own situation/program's tools
- Other suggestions for improving the intervention included:
    - Provide this material at the beginning of sites' work with WSC/part of onboarding
    - Tailor training to specific focus areas (e.g., environmental education programs)
    - Shorten the intervention ("*the process was really long"* and "*the time between the WebEx and follow up call seemed long"*)
    - Offer training as a one-day workshop held for sites in the same region, similar to what AmeriCorps members go through, which would also help share ideas and build community
    - Finding evidence—how to access research articles that align with their program

Upon reviewing participants' interview data, we noted some additional considerations/themes:

- **Role seems to matter**—in one program, a new supervisor noted multiple benefits from the intervention—but a second point of contact, who also participated but was less involved with AmeriCorps, did not feel as looped in or that it was as beneficial.
- **Status of WSC and other funding** makes a difference in the site's perception of longer term changes—One program was excited about what they learned but knew they would not be continuing with WSC so limited changes—if they were continuing with WSC, they said they would change what they track to align better with their program and new ideas. Others were concerned about how COVID-19 related closures might affect them in the year ahead.
- **Varied learning styles seem to affect how they received the intervention**—while most people felt the in-person learning with the WSC staff member was more useful than the webinars, for some the webinars were particularly tough—one person mentioned being a "kinetic learner" so needing in-person interaction. People felt the webinars contained a lot of good information, but some noted they may not necessarily meet all learning styles. Some people did the homework exercises and found they helped, and others offered that they skipped them entirely. This could be due to time/bandwidth or otherwise relate to their level of interest or learning style. At the same time, another person noted she appreciated the flexibility of being able to watch the webinars on their own time.
- **Differentiated instruction**—at least 2 programs noted that this intervention would be helpful for programs new to AmeriCorps. Others felt like they needed less time on some content but that digging into some recommendations and things about their program were helpful when meeting with WSC. Relatedly, one person noted that she just finished grad school and did an independent study that covered a lot of the content of WSC's training.

Listening Sessions

The purpose of our listening sessions was to provide a qualitative assessment of the site visit component of the intervention and the extent to which it was implemented as intended (see the full tool in **APPENDIX 4**). Some of the elements that we listened for included:

- The extent to which the site visits followed the planned agenda and topics
- The use of materials
- The level of engagement and participation by the programs
- Examples of tough questions or challenging conversations and how they were resolved
- Generalizability or the extent to which site visits were able to cover evaluation-related topics of relevance beyond specific AmeriCorps requirements
- Overall strengths and challenges of the site visits
- Issues or ideas related to replicating the intervention

Below is a summary of what we observed with respect to these different elements.

<u>Implementation Fidelity</u>

- Across the 3 sites, WSC implemented the site visit content with consistency and fidelity.
- The way the site visits were introduced varied a bit, and so did some site staff members' understanding of what the visit would entail. For instance, some challenges with WebEx meant that some meetings started later or were shorter in duration than others. In two of three site visits observed, the meetings started late due to technical difficulties, and during one visit, they continued to experience additional interruptions due to sound issues. In one case, a site had not understood that they would need a laptop or ability to view slides during the site visit, and so it took them some time to find and set up equipment. In another case, WebEx stopped working, so a site staff member needed to join by phone, and WSC ended up emailing the slides.

<u>Use of materials</u>

- WSC used materials consistently, which were customized to include site-specific examples, drawn from their application materials (e.g., the site's own logic model, evidence/research they had shared) and included WSC ideas on how they might add to or refine their work.
- Where WSC had grouped several site visits by a common focus area (e.g., environmental stewardship), they were able to use some common slides, research studies, and suggested survey tools across multiple site visits.

<u>High levels of engagement through skilled facilitation</u>

- The WSC staff member set a warm tone to the conversation and used a lot of positive language to reinforce program ideas and elements that were in good shape/aligned with WSC requirements, such as "*You're doing a good job with your reports*," "*There we go!*," and "*I think you exceeded your targets…a lot of work went into that*," as well as language that was normalizing and "joining" with the programs (e.g., "*We don't want data to be scary*" or "*In your role, you want to know that the data is telling you, right*?").
- The WSC staff member used an approach of "offering" ideas or draft thoughts/revisions for the participants to react to, and this seemed to create an effective dialogue. For example, he used phrases like, *"What I've done is I took your Theory of Change and kind of rewrote it because I like doing this stuff"* or shared examples of how he thinks about similar tasks/challenges the sites

had to do (e.g., *"[When I do] our reporting I like to look at the total who participated and <u>also</u> those who did pre and post"* and "*I'm wondering*…." and then the programs would share their ideas and questions.

- WSC and the participants exchanged ideas back and forth in a conversational way. WSC would offer a topic or return to the materials, sharing observations and feedback on the programs' own application materials (e.g., logic model, data collection tools, etc.), keeping the meeting on track as planned, but also asking the participants for their reactions to the ideas he shared. The WSC staff member used questions effectively to keep the staff engaged and to gauge their understanding of topics—almost like a playful quiz at times (e.g., *"So what are the categories of outcomes we could measure?"*; "*So let's take a look at High Quality Data. Before I go to this, without looking at your cheat sheets, can you name them?")* or to advance their thinking (e.g., *"So if you were going to revise your Theory of Change, and I think it's really good, how would you do that?"; "Does this encompass what you do?*").
- Participants engaged the WSC staff member with questions and asked for advice (e.g., about which students to survey and when, how to revise their logic model to reflect changes in who they serve or program activities, how/who to count as "participated" (using dosage vs. complete pre/post data), and how to find evidence from outside research that fit with their unique program designs).
- Participants seemed comfortable sharing feedback on the recorded sessions (e.g., *"It was pretty much review"*). Additionally, participants shared unsolicited positive comments during and at the end of the visits, such as *"…I was excited about the resources list/research with evidence,""…giving us lots of good things to think about,"* "*That's a really good idea!,"* and *"Honestly this whole thing has been beneficial. I appreciate it!"*

<u>Tough questions/challenges in the site visits</u>

- Some sites experienced technological challenges during their WebEx site visits, which may have affected their ability to attend to the content fully and created some frustration. The WSC staff member kept the agenda on track despite this and deployed work-around solutions (e.g., emailing slides when WebEx visuals could not be used).
- One challenging topic that came up more than once was about how to count how many students completed a program—do you base this on who completed a permission waiver?, how many have both pre and post outcome data points?, or how many reached a particular dosage threshold? There can be different perspectives on the best way to approach this, and when programs decide to use a criterion other than a minimum dosage threshold, they may be missing out on capturing the complete picture of how many students they serve.
- Participants had different levels of experience with research, and this seemed to affect how they engaged with the content—sometimes negatively. In one case, staff from an environmental program expressed frustration with the recorded webinars ("*So we just watched it yesterday and honestly I thought for us that it was completely unnecessary and not helpful. We are all college graduates, well-educated. It seemed like information I should teach my 13 year-old son, not me. Sorry."*). They seemed less engaged in parts of their visit (e.g., they were quiet or silent in response to WSC discussion questions) than other sites. Faced with this challenge, the WSC staff member used humor effectively and persisted in adapting the conversation well; he expressed empathy and appreciation for the expertise they had (*"I just want to elevate your documents even though the info in that webinar was generic. Here's your survey. It's valid, and you get an A+ for that," and* "*Sites don't always know how to count acreage, but we can use yours as a best practice."*). As he continued to praise their work, and moved to other topics

(e.g., sharing some research studies as evidence of their program), the site began to talk more and expressed appreciation for getting clarity on WSC expectations. Yet, this visit illustrated a common challenge in evaluation capacity-building: How do you provide TA/training on a topic that someone already knows about, or how do you press to see if they really know and understand what they think they know? One the one hand, this program noted that they were trained in environmental science and used to working with data. On the other hand, it was clear they were not knowledgeable about all topics that WSC covered (e.g., how to track participation; and policies around protecting student privacy).

- Participants had different levels of experience with WSC/CNCS requirements, so sometimes they were seeing their program's WSC logic models or other material (e.g., reporting templates) for the first time during the site visit. In all cases of this that we observed, the new supervisor or staff member expressed appreciation and described gaining a new understanding of the content and of the possibilities for enhancing their measurement and reporting in future.

<u>Generalizability</u>

- Although the site visits used content from WSC applications and fit with the WSC/CNCS requirements, a majority of the points of discussion or questions focused on universal challenges in evaluation and performance measurement, and as such were highly generalizable beyond WSC. Examples included: the right level of detail and key elements to describe in a logic model; how to balance collecting complete data with what is feasible and valid (e.g., considerations in surveying students with autism or students who cannot yet read); protecting confidentiality; simplifying metrics and focusing on some key indicators; revisiting measurement to better align tools to program activities; how to count participation or measure dosage; processes to improve data quality, such as double-checks for accuracy and ways to identify duplicates; and where to find supporting evidence and research from other programs.

<u>Overall strengths and challenges and implications for scaling</u>

- Many of the strengths are highlighted above, including consistent use of materials, customization/tailoring of content to specific programs, and strong facilitation skills.
- Reviewing the purpose of the site visits and sharing the planned agenda at the start of the WebEx may help participants focus and feel more comfortable, in case they missed prior messaging.
- Each program had one or more challenges solved on the spot or "a-ha" moments during the site visits, which they appreciated and revealed the effectiveness of the visit. These included: learning about articles/research evidence that they could use to support their program; concrete revisions to improve their Theories of Change or Logic Models; streamlining the number of performance measures they used or selecting more valid methods (e.g., counting the number of clean ups done by kayak rather than acreage); concrete ideas on modifications to the content and/or timing of surveys; ideas for future evaluation studies; and ways to more accurately analyze and report their results.
- It was evident that the WSC staff member knew the programs' histories and their staff well, so he was able to share connections and best practices between similar programs, bring up ways they had changed or improved their design and measurement over time, and reference details of their work to illustrate concepts.

Recommendation Reports

After each site visit by WebEx, WSC prepared and sent a thank you email, which included a Recommendation Report, an electronic copy of the PowerPoint deck used during the visit, and additional resources (e.g., examples of relevant surveys, curricula, or research studies). Due to the added responsibilities of scheduling and other logistics, WSC's staff member had a temporary backlog of recommendation report writing, so WSC sent the reports and held the follow up meetings with sites slightly later than originally intended. Generally, WSC sent recommendation reports and held follow up calls within 6-8 weeks of site visits.

Recommendation reports included specific, detailed recommendations grouped in categories that matched the recorded session topics and the structure of the site visit: Program Alignment, Evidence of Effectiveness, Data Collection and Use, and High-Quality Data.

The evaluation team reviewed 7 recommendation reports and follow up materials in detail, focusing on sites who had consented to listening sessions and interviews. In the case of four of these sites, a member of the evaluation team had listened to the sessions and conversations that had fed into the recommendation reports. The content of the recommendation reports matched well with what we observed in the listening sessions.

In reviewing the recommendation reports (and accompanying materials) we observed:

- **A high level of detail/specificity and customization.** WSC provided specific, actionable feedback/recommendations to the sites for consideration, aligned with the content presented via the recorded sessions. This helped tangibly connect concepts from the recorded sessions to the site's own materials and program so they could see how the material could be applied
- **Recommendations extended and reinforced site visit conversations.** Recommendation reports included ideas and challenges from conversations during the site visits. Subheadings grouped the recommendations into categories aligned with webinar content, and succinct bullets summarized ideas digestibly.
- **All programs received feedback in all areas, but some programs received more recommendations than others.** Every recommendation report offered ideas or feedback across the 4 areas of evaluation and performance measurement covered by the intervention. When a site seemed to excel in a particular area, the recommendation report shared positive feedback and examples reinforcing that. Other programs received concrete recommendations across multiple areas ranging from suggested measurement tools, considering conducting new evaluation studies, revising logic models, or refining data collection or reporting processes.
- **Where relevant, sites received new resources such as published studies or examples of measurement tools or logic models.** WSC shared electronic versions of materials they shared during the site visits and in some cases augmented them with additional materials, customized to each program.
- **Common challenges appeared across multiple sites/programs.** Although some programs received more or fewer recommendations than others and had their own contexts, backgrounds in research, and understanding of WSC/AmeriCorps requirements, some themes appeared across multiple reports, and they were consistent with what we observed in the pilot year of this study. Common challenges included:

- **Content and format of Theories of Change and Logic models**. All 7 sites received suggestions on the right level of detail to include in a Theory of Change and Logic Model, such as noting what was missing from theirs (e.g., including needs addressed, inputs/resources specific to the program, intended dosage, numeric targets, descriptions of core program activities, etc.) or suggesting ways to streamline or consolidate information that had been included (e.g., one site had used 3 different logic models to describe their program, and another had included details not relevant to the program). Some sites also received guidance around formatting, such as ways to show alignment across activities, outputs, and outcomes by arranging them in a linear fashion, or on making sure that descriptions of outcomes or other elements were consistent across their Theory of Change and Logic Model.
- **Reviewing, updating, searching for literature/published evidence to provide evidence of effectiveness**. Among 6 of 7 sites, WSC shared research studies that were new to them, which helped to support their program and in some cases provided ideas about new innovations to their program design or data collection methods (e.g., WSC shared a meta-analysis on commitment-making and the use of pledges as an effective element of environmental programs, with several programs doing this work.
- **Refining data collection methods.** Several programs received recommendations regarding their data collection methods, such as survey tools to consider, clarifying which beneficiaries to survey and when, considering changes to the outcomes measured when sites had shifted their focus over time or learned with experience that a measure did not make sense (e.g., not counting acres when doing kayak cleanups on the water), adding new measures (e.g., starting to measure behavior change), or streamlining the number of tools they used).
- **Data quality/Reporting**. WSC clarified their expectations and in some cases reviewed templates for how sites should summarize their data in quarterly reports to WSC. In several cases, WSC shared tangible recommendations about how to improve data quality that extended beyond the context of WSC/CNCS, such as reminding a program not to share student identifiers and suggesting ways to tighten up processes of checking the accuracy of data.

Staff Interview

Our staff interview revealed consistency with our observations and also provided a window into some of the factors and challenges for WSC or others to consider in replicating this intervention. Some of the key themes are presented below:

<u>Successes and related factors</u>

- Overall, the intervention felt successful and largely followed the planned timeline. Things went smoothly in terms of scheduling and meeting with sites.
- Piloting the intervention in year 1 was important
    - It helped the staff member build confidence in delivering it;
    - Narrowing down the content of the webinars in response to year 1 feedback helped improve it; and
    - The pilot also helped WSC leadership, who had been new in year 1, understand it better this year, and they were able to provide more support and involvement.

- Early weekly communications and meetings between members of the WSC team (and our evaluation team) helped to keep things get started and moving on track, even with an accelerated timeline.
- Grouping the programs in cohorts/scheduling by common focus areas helped WSC be more efficient and use some of the same slides and studies to prepare (e.g., when meeting with multiple environmental programs), because they had some of the same or similar program design elements, activities, and evidence base. Typically, the staff member could prepare 1.5 slide decks per day in preparation for site visits.
- Supervisors tended to stay engaged in the intervention, particularly new supervisors.
- In general, new staff seemed engaged, eager to learn, and they were able to identify some problems in their data collection or reporting when meeting with WSC, which led to tangible improvements. They shared how the intervention helped them better understand WSC's expectations around evidence, evaluation, reporting, and how to more effectively meet those requirements.
- The relationship between the WSC staff member and the sites—knowing their culture and layout, knowing their roles and experience—helped make the intervention more effective.
- Environmental stewardship programs seemed to demonstrate a strong commitment to the intervention, and WSC felt this may be because their staff tended to have a background/training in science.
- Content on Evidence, Data Collection and Use, and Data Quality seemed to be the most successful. Sites appreciated learning about other studies and how to describe them as evidence in relation to their own program. Some sites were further along than others in their data collection and data quality efforts, but overall, many were able to identify some issues, get a sense of the bigger pictures, and see how their data related to other programs and WSC.

<u>Challenges</u>

- Support staffing for the intervention to handle logistics changed. Although WSC had originally planned on providing administrative support to help with scheduling site visits and other logistics, in the end, Robert took on these tasks. This affected how quickly Robert could write recommendation reports and schedule follow up visits, although he was able to keep the overall intervention schedule on time.
- Notes from the site visits varied, which added time and complexity to writing recommendation reports. The original plan called for a consistent WSC staff member to serve as note-taker during site visits. This plan also changed, and with multiple note-takers, who were not necessarily knowledgeable about the intervention content, notes varied. Robert observed in both years of this study that having a second staff member, who is familiar with the content, serve as a note taker makes writing recommendation reports more efficient and effective. A note taker who *"understands the right language"* is key.
- Technological challenges with WebEx affected some site visits; because some sites experienced difficulties connecting to WebEx, or they had not realized in advance that they would need to connect for the site visits, some visits started late. WSC also had changed versions of WebEx and was unable to record site visits as originally planned.
- Content around program alignment felt the least successful and the least understood, in part because some programs may have their own Theory of Change, apart from their work with WSC, or just may not be as focused on it.
- Engaging leadership from the sites' programs was challenging. WSC aimed to engage staff at the Supervisor level at a minimum but getting others was difficult.

- Recruitment and retention of AmeriCorps members was challenging this year, and WSC lost some sites as a result. Next year, they may be dropping from 500 to 350 AmeriCorps members. This and the COVID-19 pandemic added to challenges in scheduling and in future plans for this intervention.

Suggestions for replication and scaling

- A Learning Management System (LMS) can centralize how intervention materials are shared (e.g., webinars), offer flexibility for participants to view and revisit information in smaller chunks, and affords opportunities to track who participates and collect other data. Next year, WSC will have an LMS in place for AmeriCorps members and staff.
- Data collections plans were helpful to have in place as well as looking at how individual sites used data (e.g., by reviewing their websites, reviewing their reporting to WSC). This helped make discussions concrete and actionable.
- Aligned measures are important topic for new sites, which they may not know about, so this may be beneficial to cover as they start with WSC/AmeriCorps.
- Smaller AmeriCorps programs may be able to replicate this intervention well, perhaps working with one organization at a time.
- Although most sites seemed to benefit from the intervention, WSC could envision offering some tiered learnings/offerings in the future (e.g., Evaluation 101, 201).

APPENDIX 1: WSC Training and Technical Assistance Intervention – Original Program Logic Model

Inputs	Activities	Output Targets	Short-Term Outcomes	Mid-Term Outcomes	Long-Term Outcomes
Training Curriculum/Intervention WSC Intervention delivery staff *(i.e., program manager and 3 core staff trained to deliver curriculum to sites)* Program evaluation consultant and team Participant Sites / Individual staff participants at the sites Training materials Original proposal/application materials "Technology" – systems for delivering successful webinars, phone and WebEx "on-site" visits Evaluation Design to assess implementation and impact Data Collection tools to gather data on implementation and impact	**WSC Staff Delivery of the Training Intervention and Technical Assistance** *Recorded training sessions* in substantive areas: Program Alignment, Evidence of Effectiveness, Data Collection and Use and High-Quality Data *Skill Building Activities and Skills Checklists* *Individualized "site visit"* either on-site, via WebEx or via phone *Recommendations Report Follow-up Technical Assistance sessions* **Evaluation of program implementation and impact** Collection of Pre and Post Survey data from participants in the intervention Interviews with selected staff from participating sites Collection of intervention implementation data from staff delivering the training	**Delivery of training intervention to 18 participating sites across 3 "cohorts"** 3 WSC Training Staff deliver intervention to 2 participating sites in each of the cohorts Delivery of all intervention components to a site over a 2-3-month time period Estimated 1-3 participants per site - estimated 30-50 total participants across all sites Cohort A: participation in five recorded webinars Cohort B: participation in webinars, 1 WebEx "site visit", recommendation report and one phone TA session Cohort C: participation in webinars, 1 in-person "site visit", recommendation report and one phone TA session **Evaluation of Program** 30-50 Total participants complete surveys 1 month before and after intervention About 10-12 participants complete interviews 1-2 months after the training intervention 3 core training staff complete implementation assessments	**Increased participant knowledge related to:** Program alignment (performance measures), Evidence of effectiveness, Data collection and use, Data Collection Strategy and, High quality data **Increased participant beliefs in the importance of _____ in providing high quality program services:** defining community needs and aligning all program elements to address needs, developing and using a program logic model, using evidence of effectiveness when selecting or designing program interventions, writing and implementing a high-quality data collection plan, using data to inform decision-making about their program, and collecting high quality performance measurement data	**Increased participant ability to implement components of performance measurement and evaluation process including:** defining community needs and aligning all program elements to address needs, developing and using a program logic model, using evidence of effectiveness when selecting or designing program interventions, writing and implementing a high-quality data collection plan, using data to inform decision-making about their program, collecting high quality performance measurement data **Program implementing components of performance measurement and evaluation process with greater quality and frequency:** defining community needs and aligning all program elements to address needs, developing and using a program logic model, using evidence of effectiveness when selecting or designing program interventions, writing and implementing a high-quality data collection plan, using data to inform decision-making about their program, and collecting high quality performance measurement data	Build further evidence Use Data and Findings Deliver Stronger Programs Build further evidence for supporting program efforts See Better results for program participants

APPENDIX 2: WSC Evaluation BASELINE Survey

Thank you for your willingness to complete this survey. This baseline survey gathers information about your knowledge and attitudes related to program evaluation and performance measurement. We will ask you to complete this survey now and then a similar version in about 4 months. Parts of the survey may feel like a "quiz," asking questions about specific concepts. This is to help our research team understand the extent to which individuals' knowledge of these concepts change over time, not to test or make judgments about your answers.

There are about 30 questions in the survey, and it should take about 15-20 minutes. Please be as honest as possible with your answers. The information on the survey is confidential, and only research staff will have access to your individual responses. You will complete an ID code question at the start that will help us track responses across the two survey time periods. If you have any questions, please contact Marc Bolan at XXXXXX

1. What is your current organization/agency? _____
2. How long have you worked in your current role with the organization? _____ YEARS _____ MONTHS

3. What is your current job title? _____

4. Which the following best describes your role in working with the AmeriCorps program and the Washington Service Corps? (*choose all that apply*)

- ❑ Primary Site Staff
- ❑ Secondary Site Staff
- ❑ Quarterly Reporting Staff
- ❑ Other _____

5. How long have you worked with a site supported by the Washington Service Corps?
____ YEARS ____MONTHS
6. How long have you worked as an AmeriCorps Supervisor at the current organization?
___ YEARS __ MONTHS
7. How long have you worked with or otherwise been associated with AmeriCorps programs in any capacity?
____ YEARS _____ MONTHS
8. Have you ever served as an AmeriCorps member? YES NO

9. How would you best describe your experience with program evaluation, performance measurement, and/or data collection activities? *(choose all that apply)*

- ❑ I have no or very little experience with any of these activities
- ❑ I have some experience with program evaluation
- ❑ I have some experience with AmeriCorps performance measures and collecting data for these measures
- ❑ I have been involved in all these activities for many years and have expertise in carrying them out!

10. When we speak of the evidence of effectiveness of an intervention we refer to *(choose one answer)*

- ❑ Studies indicating that a program is likely to lead to a specific effect or outcome
- ❑ Studies indicating that program staff believe that the intervention is working
- ❑ Studies indicating that a program carried out the intervention as intended
- ❑ Studies indicating that a program met the expectations of the funding agency

11. Which of the following is NOT true about the process of performance measurement? *(choose one answer)*

- ❑ It involves the collection of output and outcome data
- ❑ It measures performance against targets
- ❑ It gathers data that is different from the data a program uses to inform decision making
- ❑ It allows the user to test personal observations with empirical data

12. You are delivering a parenting workshop to a group of participants over a three-month period. Look at each statement and choose the part of a logic model where it would best fit:

	Input	Activity	Output	Outcome
a) 40 parents complete eight weekly sessions	❑	❑	❑	❑
b) 30 parents report increased knowledge of support services available to their family	❑	❑	❑	❑
c) The staff present information on positive discipline techniques	❑	❑	❑	❑
d) 3 full time staff manage and deliver the program	❑	❑	❑	❑
e) 50 teen parents increase their use of positive discipline techniques	❑	❑	❑	❑

13. It is important to use data to inform decision making about the program. If your results show that you do not meet your outcome targets, some appropriate next steps could be *(choose all that apply)*

- ❑ Increase program outreach to other target populations
- ❑ Expand the same services to more participants
- ❑ Consider finding another evidence-based practice to try and replicate
- ❑ Find a new way to deliver similar services
- ❑ Consider ending the program services
- ❑ Completely adjust your logic model to fit your results

14. Your program offers environmental education workshops for high school youth helping them understand stewardship practices to save water and energy. You are interested in determining whether their knowledge of these practices increases over time. Which of the following would be a VALID method for measuring this outcome: *(choose all that apply)*

- ❑ One survey completed at the end of the workshop asking them their current knowledge of the practices
- ❑ A pre-post survey testing knowledge of these practices before and after the workshop
- ❑ Site visits to observe youth trying the practices before and after the workshop
- ❑ An interview with selected participants at the end of the workshop asking about specific elements of the practices they know more about after the workshop
- ❑ A tracking sheet indicating whether the youth attended a specific workshop

15. Which of the following would be a concern with the <u>consistency</u> of the data *(choose one answer)*?

- ❑ The program uses one survey to measure program outputs and a second survey to measure program outcomes
- ❑ The program only collects data from one of the four sites involved in the intervention
- ❑ The staff use an observational tool to measure changes in client knowledge
- ❑ Staff at one location use one survey to measure outcomes; staff at second location use a different survey to measure the same outcomes

16. Which of the following is the best example of a thorough intervention statement? *(choose one answer)*

- ❑ Our program distributes informational materials to interested community members
- ❑ Over 200 of the students in the 5 different after school tutoring programs demonstrate increased academic performance over the school year
- ❑ The program provides substance abuse prevention training to 300 middle school youth through 8 2-hour weekly sessions in 5 sites
- ❑ Our program gives out weekly vouchers to homeless families

17. When building your program's evidence of effectiveness …. (*choose one answer*)

- ❑ past performance measurement data on outcomes cannot be used as evidence
- ❑ demonstrating positive change on an outcome pre to post may count as preliminary evidence
- ❑ a program's life cycle determines the standard for preliminary evidence

- ❑ results from studies of other programs should not be considered

18. Your program has an outcome that participants in an Earth Day event increase their use of environmentally practices after the event. Which of the following below would be the best way to collect data to inform you about that outcome? *(choose one answer)*

- ❑ Observation of the actions of individuals at a future event
- ❑ A focus group three months after the event asking individuals what they liked about the event
- ❑ A self-report survey completed 3 months after the event asking them what they know about different practices
- ❑ An interview 2 months after the event asking them about their daily conservation practices

19. Which of the following best describes higher levels of evidence on the continuum of Evidence of Effectiveness *(choose one answer)*

- ❑ Research studies at higher levels of evidence have larger sample sizes
- ❑ Research studies at higher levels of evidence look at behavioral rather than knowledge or attitudinal outcomes
- ❑ Research studies at higher levels of evidence include a comparison or control group
- ❑ Research studies at higher levels of evidence do not include qualitative data.

20. When reviewing some performance measure results, we find that the number of individuals who achieved an outcome well exceeded the number of individuals participating in the program. In this case, we are most concerned with: *(choose one answer)*

- ❑ The validity of the results
- ❑ The accuracy of data collection and analysis
- ❑ The consistency of the data
- ❑ The generalizability of the results

21. Which of the following are the best examples of a study that will meet the standard of preliminary evidence for an outdoor science education program for middle schoolers (*choose all that apply*)

- ❑ A study on how the program was implemented
- ❑ A study demonstrating that 85 middle schoolers enjoyed the program
- ❑ A study demonstrating greater knowledge of science concepts among program middle schoolers than a control group in the spring
- ❑ A study demonstrating increased knowledge among program middle schoolers from fall to spring

22. A change in how someone felt about something would be an example of an *(choose one answer)*:

- ❑ Conditional outcome
- ❑ Knowledge output
- ❑ Attitudinal outcome
- ❑ Behavioral demonstration
- ❑ Attitudinal input

23. Thinking about the <u>previous question 22</u>, which of the following would be appropriate ways of measuring whether that change occurred *(choose all that apply)*

- ❑ Pre/Post Knowledge Assessment
- ❑ Observation of the individual's actions
- ❑ Open-ended interview with individual
- ❑ Pre/Post Self-report survey
- ❑ Review of a case manager's case notes about things worked on with the individual

24. A program provides financial planning workshops to low income families. Which statements below represent outcomes that align with the program activity (*choose all that apply*)

- ❑ 80 participants increase their knowledge of how to set up a savings account
- ❑ 70 percent of participants who enroll are from low-income households
- ❑ 100 individuals attend workshops in three different locations
- ❑ 200 participants demonstrate the skills needed to balance personal budgets
- ❑ 200 participants demonstrate improved attitudes toward parenting
- ❑ 5 staff train 3 AmeriCorps members in how to deliver the program

25. Which of the following measures would inform you of whether your program has implemented program services as intended *(choose all that apply)*

- ❑ The number of clients who would refer their friends to your program
- ❑ The dosage of services provided to each client
- ❑ Each client's level of gain in knowledge
- ❑ A staff member's willingness to provide more services for the agency
- ❑ The frequency of delivery of the program activities

26. A Job Search Training program that serves adults drafted a logic model. Take a look and help us correct some mistakes in it.

Inputs	Activity	Output	Short term outcome	Long term outcome
3 AmeriCorps members Training kits and materials 8 staff members 3 community partners 100 adults complete job training	Resume writing assistance Mock interviews Job shadowing	100 adults complete job search training 80 adults demonstrate new job search skills as measured by a pre—post assessment	60 of the adults trained stay in their jobs for at least 5 years 80 adults demonstrate new job search skills by the end of the program in a pre-post assessment 150 adults increase their self-confidence in their job search skills by the end of the program as measured by a pre/post survey	The employment rate in the community served increases across 5 years

26. You are assessing the quality of the logic model above. Which of the following statements are correct? *(choose all that apply)*

- ❑ "100 adults complete job search training" SHOULD NOT be an input
- ❑ "100 adults complete job search training" SHOULD NOT be an output
- ❑ "80 adults demonstrate new job search skills" SHOULD NOT be an output
- ❑ "150 adults increase their self-confidence in their job search skills" is NOT APPROPRIATE because self-confidence is not a possible outcome.
- ❑ "150 adults increase their self-confidence in their job search skills" is NOT APPROPRIATE because a change for 150 is not a possible outcome
- ❑ The outcomes DO NOT ALIGN with the program design
- ❑ Not all of the short-term outcomes are short term
- ❑ A 5-year outcome should not be in a logic model

THE NEXT SET OF QUESTIONS ASK ABOUT YOUR OWN ATTITUDES AND/OR BELIEFS RELATED TO PROGRAM EVALUATION AND PERFORMANCE MEASUREMENT

28. Please assess your level of agreement with the following statements	**Strongly Agree**	**Slightly Agree**	**Neither Agree nor Disagree**	**Slightly Disagree**	**Strongly Disagree**
Performance measures, or setting goals and measuring progress with data, are a necessary component of our work	❑	❑	❑	❑	❑
We should use evaluation data in the planning and implementation of program services	❑	❑	❑	❑	❑
If we had more time and money, I would invest it in improving how we collect and use data	❑	❑	❑	❑	❑
We can make necessary decisions about our program services with little or no data	❑	❑	❑	❑	❑
Data about program implementation and outcomes strongly influences our planning for future services	❑	❑	❑	❑	❑
We should draw from research studies and evidence when developing and planning program services	❑	❑	❑	❑	❑
The data we gather from participants is only needed for reporting to our funders and interested audiences	❑	❑	❑	❑	❑

29. "How would you rate <u>your confidence in your own ability</u> to…."	**Cannot do at all**	**Can do a little**	**Can do moderately well**	**Can do very well**
Create a program logic model	❑	❑	❑	❑
Gather information to build evidence for our program	❑	❑	❑	❑
Specify clear and detailed program outputs, outcomes and performance measures	❑	❑	❑	❑
Collect data about program performance and outcomes	❑	❑	❑	❑
Design or identify appropriate data collection tools for the right situation	❑	❑	❑	❑
Develop a detailed plan for collecting program data	❑	❑	❑	❑
Develop or identify a formal method to store and/or manage program data	❑	❑	❑	❑
Analyze and interpret program performance and outcome data	❑	❑	❑	❑
Assess the validity, reliability and consistency of the data collected	❑	❑	❑	❑
Identify/use evidence/data to establish the needs of the community	❑	❑	❑	❑
Use data to inform decision making about the program	❑	❑	❑	❑

THE NEXT SET OF QUESTIONS ASK ABOUT YOUR PERCEPTIONS OF PROGRAM EVALUATION AND PERFORMANCE MEASUREMENT AT YOUR PROGRAM/ORGANIZATION/AGENCY. IT IS OK IF YOU ARE NOT SURE ABOUT WHAT YOUR ORGANIZATION HAS DONE

30. To what extent has our <u>program/organization/agency</u> done the following activities:	**No, we have not done**	**Yes, we've done some of this, but we could do much better**	**Yes, and we have done a pretty good job with this**	**Yes, and we have done this very well**	**I don't know**
Created a complete and detailed logic model that reflects our program	❑	❑	❑	❑	❑
Used and reviewed our logic model to support decision making and program delivery	❑	❑	❑	❑	❑
Created a detailed plan for collecting output, outcome and performance measure data	❑	❑	❑	❑	❑
Developed a plan or strategy for checking the quality of the data that is collected	❑	❑	❑	❑	❑
Developed or identified a formal method or system to store and/or manage program data	❑	❑	❑	❑	❑
Use data to inform decision making about the program and/or organization as a whole	❑	❑	❑	❑	❑
Created a plan to review program data consistently	❑	❑	❑	❑	❑
Trained program staff on how to collect and use outcome and performance data	❑	❑	❑	❑	❑
Assigned data collection and evaluation functions to specific people *(e.g., who collects data, who enters it, who reviews it, etc.)*	❑	❑	❑	❑	❑

APPENDIX 3: WSC Evaluation FOLLOW-UP Survey

WSC Evaluation Follow-Up Survey

Thank you for your willingness to complete this survey. This survey gathers information about your knowledge and attitudes related to program evaluation and performance measurement. You completed a similar survey about 4 months ago! Parts of the survey may feel like a "quiz," asking questions about specific concepts. This is to help our research team understand the extent to which individuals' knowledge of these concepts change over time, not to test or make judgments about your answers.

There are about 30 questions in the survey, and it should take about 15-20 minutes. Please be as honest as possible with your answers. The information on the survey is confidential, and only research staff will have access to your individual responses. You will complete an ID code question at the start that will help us track responses across the two survey time periods. If you have any questions, please contact Marc Bolan at XXXX

1. What is your current organization/agency? _____
2. How long have you worked in your current role with the organization? _____ YEARS _____ MONTHS

3. Which the following best describes your role in working with the AmeriCorps program and the Washington Service Corps? (*choose all that apply*)

- ❑ Primary Site Staff
- ❑ Secondary Site Staff
- ❑ Quarterly Reporting Staff
- ❑ Other _____

4. A change in how someone felt about something would be an example of an *(choose one answer)*:

- ❑ Conditional outcome
- ❑ Knowledge output
- ❑ Attitudinal outcome
- ❑ Behavioral demonstration
- ❑ Attitudinal input

5. Thinking about the previous question 22, which of the following would be appropriate ways of measuring whether that change occurred *(choose all that apply)*

- ❑ Pre/Post Knowledge Assessment
- ❑ Observation of the individual's actions
- ❑ Open-ended interview with individual
- ❑ Pre/Post Self-report survey
- ❑ Review of a case manager's case notes about things worked on with the individual

6. Your program offers environmental education workshops for high school youth helping them understand stewardship practices to save water and energy. You are interested in determining whether their knowledge of these practices increases over time. Which of the following would be a VALID method for measuring this outcome: *(choose all that apply)*

- ❑ One survey completed at the end of the workshop asking them their current knowledge of the practices
- ❑ A pre-post survey testing knowledge of these practices before and after the workshop
- ❑ Site visits to observe youth trying the practices before and after the workshop
- ❑ An interview with selected participants at the end of the workshop asking about specific elements of the practices they know more about after the workshop
- ❑ A tracking sheet indicating whether the youth attended a specific workshop

7. When reviewing some performance measure results, we find that the number of individuals who achieved an outcome well exceeded the number of individuals participating in the program. In this case, we are most concerned with: *(choose one answer)*

- ❑ The validity of the results
- ❑ The accuracy of data collection and analysis
- ❑ The consistency of the data
- ❑ The generalizability of the results

8. Which of the following is NOT true about the process of performance measurement? *(choose one answer)*

- ❑ It involves the collection of output and outcome data
- ❑ It measures performance against targets
- ❑ It gathers data that is different from the data a program uses to inform decision making
- ❑ It allows the user to test personal observations with empirical data

9. Which of the following measures would inform you of whether your program has implemented program services as intended *(choose all that apply)*

- ❑ The number of clients who would refer their friends to your program
- ❑ The dosage of services provided to each client
- ❑ Each client's level of gain in knowledge
- ❑ A staff member's willingness to provide more services for the agency
- ❑ The frequency of delivery of the program activities

10. It is important to use data to inform decision making about the program. If your results show that you do not meet your outcome targets, some appropriate next steps could be *(choose all that apply)*

- ❑ Increase program outreach to other target populations
- ❑ Expand the same services to more participants
- ❑ Consider finding another evidence-based practice to try and replicate
- ❑ Find a new way to deliver similar services
- ❑ Consider ending the program services
- ❑ Completely adjust your logic model to fit your results

11. You are delivering a parenting workshop to a group of participants over a three-month period. Look at each statement and choose the part of a logic model where it would best fit:

	Input	Activity	Output	Outcome
a) 40 parents complete eight weekly sessions	❑	❑	❑	❑
b) 30 parents report increased knowledge of support services available to their family	❑	❑	❑	❑
c) The staff present information on positive discipline techniques	❑	❑	❑	❑
d) 3 full time staff manage and deliver the program	❑	❑	❑	❑
e) 50 teen parents increase their use of positive discipline techniques	❑	❑	❑	❑

12. Which of the following is the best example of a thorough intervention statement? *(choose one answer)*

- ❑ Our program distributes informational materials to interested community members
- ❑ Over 200 of the students in the 5 different after school tutoring programs demonstrate increased academic performance over the school year
- ❑ The program provides substance abuse prevention training to 300 middle school youth through 8 2-hour weekly sessions in 5 sites
- ❑ Our program gives out weekly vouchers to homeless families

13. Your program has an outcome that participants in an Earth Day event increase their use of environmentally practices after the event. Which of the following below would be the best way to collect data to inform you about that outcome? *(choose one answer)*

- ❑ Observation of the actions of individuals at a future event
- ❑ A focus group three months after the event asking individuals what they liked about the event
- ❑ A self-report survey completed 3 months after the event asking them what they know about different practices
- ❑ An interview 2 months after the event asking them about their daily conservation practices

14. Which of the following would be a concern with the <u>consistency</u> of the data *(choose one answer)*?

- ❑ The program uses one survey to measure program outputs and a second survey to measure program outcomes
- ❑ The program only collects data from one of the four sites involved in the intervention
- ❑ The staff use an observational tool to measure changes in client knowledge
- ❑ Staff at one location use one survey to measure outcomes; staff at second location use a different survey to measure the same outcomes

15. Which of the following best describes higher levels of evidence on the continuum of Evidence of Effectiveness *(choose one answer)*

- ❑ Research studies at higher levels of evidence have larger sample sizes
- ❑ Research studies at higher levels of evidence look at behavioral rather than knowledge or attitudinal outcomes
- ❑ Research studies at higher levels of evidence include a comparison or control group
- ❑ Research studies at higher levels of evidence do not include qualitative data.

16. Which of the following are the best examples of a study that will meet the standard of preliminary evidence for an outdoor science education program for middle schoolers (*choose all that apply*)

- ❑ A study on how the program was implemented
- ❑ A study demonstrating that 85 middle schoolers enjoyed the program
- ❑ A study demonstrating greater knowledge of science concepts among program middle schoolers than a control group in the spring
- ❑ A study demonstrating increased knowledge among program middle schoolers from fall to spring

17. When we speak of the evidence of effectiveness of an intervention we refer to *(choose one answer)*

- ❑ Studies indicating that a program is likely to lead to a specific effect or outcome
- ❑ Studies indicating that program staff believe that the intervention is working
- ❑ Studies indicating that a program carried out the intervention as intended
- ❑ Studies indicating that a program met the expectations of the funding agency

18. A program provides financial planning workshops to low income families. Which statements below represent outcomes that align with the program activity (*choose all that apply*)

- ❑ 80 participants increase their knowledge of how to set up a savings account
- ❑ 70 percent of participants who enroll are from low-income households
- ❑ 100 individuals attend workshops in three different locations
- ❑ 200 participants demonstrate the skills needed to balance personal budgets
- ❑ 200 participants demonstrate improved attitudes toward parenting
- ❑ 5 staff train 3 AmeriCorps members in how to deliver the program

19. When building your program's evidence of effectiveness …. (*choose one answer*)

- ❑ past performance measurement data on outcomes cannot be used as evidence
- ❑ demonstrating positive change on an outcome pre to post may count as preliminary evidence

- ❑ a program's life cycle determines the standard for preliminary evidence
- ❑ results from studies of other programs should not be considered

20. A Job Search Training program that serves adults drafted a logic model. Take a look and help us correct some mistakes in it.

Inputs	Activity	Output	Short term outcome	Long term outcome
3 AmeriCorps members Training kits and materials 8 staff members 3 community partners 100 adults complete job training	Resume writing assistance Mock interviews Job shadowing	100 adults complete job search training 80 adults demonstrate new job search skills as measured by a pre—post assessment	60 of the adults trained stay in their jobs for at least 5 years 80 adults demonstrate new job search skills by the end of the program in a pre-post assessment 150 adults increase their self-confidence in their job search skills by the end of the program as measured by a pre/post survey	The employment rate in the community served increases across 5 years

21. You are assessing the quality of the logic model above. Which of the following statements are correct? ***(choose all that apply)***

- ❑ "100 adults complete job search training" SHOULD NOT be an input
- ❑ "100 adults complete job search training" SHOULD NOT be an output
- ❑ "80 adults demonstrate new job search skills" SHOULD NOT be an output
- ❑ "150 adults increase their self-confidence in their job search skills" is NOT APPROPRIATE because self-confidence is not a possible outcome.
- ❑ "150 adults increase their self-confidence in their job search skills" is NOT APPROPRIATE because a change for 150 is not a possible outcome
- ❑ The outcomes DO NOT ALIGN with the program design
- ❑ Not all of the short-term outcomes are short term
- ❑ A 5-year outcome should not be in a logic model

THE NEXT SET OF QUESTIONS ASK ABOUT YOUR OWN ATTITUDES AND/OR BELIEFS RELATED TO PROGRAM EVALUATION AND PERFORMANCE MEASUREMENT

22. Please assess your level of agreement with the following statements	**Strongly Agree**	**Slightly Agree**	**Neither Agree nor Disagree**	**Slightly Disagree**	**Strongly Disagree**
Performance measures, or setting goals and measuring progress with data, are a necessary component of our work	❑	❑	❑	❑	❑
We should use evaluation data in the planning and implementation of program services	❑	❑	❑	❑	❑
If we had more time and money, I would invest it in improving how we collect and use data	❑	❑	❑	❑	❑
We can make necessary decisions about our program services with little or no data	❑	❑	❑	❑	❑
Data about program implementation and outcomes strongly influences our planning for future services	❑	❑	❑	❑	❑
We should draw from research studies and evidence when developing and planning program services	❑	❑	❑	❑	❑
The data we gather from participants is only needed for reporting to our funders and interested audiences	❑	❑	❑	❑	❑

23. "How would you rate <u>your confidence in your own ability</u> to…."	**Cannot do at all**	**Can do a little**	**Can do moderately well**	**Can do very well**
Create a program logic model	❑	❑	❑	❑
Gather information to build evidence for our program	❑	❑	❑	❑
Specify clear and detailed program outputs, outcomes and performance measures	❑	❑	❑	❑
Collect data about program performance and outcomes	❑	❑	❑	❑
Design or identify appropriate data collection tools for the right situation	❑	❑	❑	❑
Develop a detailed plan for collecting program data	❑	❑	❑	❑
Develop or identify a formal method to store and/or manage program data	❑	❑	❑	❑
Analyze and interpret program performance and outcome data	❑	❑	❑	❑
Assess the validity, reliability and consistency of the data collected	❑	❑	❑	❑
Identify/use evidence/data to establish the needs of the community	❑	❑	❑	❑
Use data to inform decision making about the program	❑	❑	❑	❑

THE NEXT SET OF QUESTIONS ASK ABOUT YOUR PERCEPTIONS OF PROGRAM EVALUATION AND PERFORMANCE MEASUREMENT AT YOUR PROGRAM/ORGANIZATION/AGENCY. IT IS OK IF YOU ARE NOT SURE ABOUT WHAT YOUR ORGANIZATION HAS DONE

24. To what extent has our <u>program/organization/agency</u> done the following activities:	**No, we have not done**	**Yes, we've done some of this, but we could do much better**	**Yes, and we have done a pretty good job with this**	**Yes, and we have done this very well**	**I don't know**
Created a complete and detailed logic model that reflects our program	❑	❑	❑	❑	❑
Used and reviewed our logic model to support decision making and program delivery	❑	❑	❑	❑	❑
Created a detailed plan for collecting output, outcome and performance measure data	❑	❑	❑	❑	❑
Developed a plan or strategy for checking the quality of the data that is collected	❑	❑	❑	❑	❑
Developed or identified a formal method or system to store and/or manage program data	❑	❑	❑	❑	❑
Use data to inform decision making about the program and/or organization as a whole	❑	❑	❑	❑	❑
Created a plan to review program data consistently	❑	❑	❑	❑	❑
Trained program staff on how to collect and use outcome and performance data	❑	❑	❑	❑	❑
Assigned data collection and evaluation functions to specific people *(e.g., who collects data, who enters it, who reviews it, etc.)*	❑	❑	❑	❑	❑

CONTROL GROUP FEEDBACK QUESTIONS

In the past 4 months, did you participate in any trainings or technical assistance that covered program evaluation/performance measurement? YES/NO

In the past 4 months, did your agency participate in any additional trainings or technical assistance that covered program evaluation/performance measurement?

(if YES to either) Please describe who provided the training/technical assistance and what topics it covered.

TREATMENT GROUP FEEDBACK QUESTIONS

THE FOLLOWING QUESTIONS ASK ABOUT YOUR PARTICIPATION IN THE WSC TRAINING AND TECHNICAL ASSISTANCE.

"Please assess your level of agreement with the following statements"	**Strongly Agree**	**Slightly Agree**	**Slightly Disagree**	**Strongly Disagree**
It was easy and convenient to participate in the training	❑	❑	❑	❑
The WSC staff was effective in communicating the content during the recorded training sessions	❑	❑	❑	❑
The WSC staff was effective in communicating the needs and requirements associated with this training	❑	❑	❑	❑
The WSC staff was responsive to our questions and/or concerns about participation in the training	❑	❑	❑	❑

From the WSC Recorded Training sessions, what was <u>most useful</u> in helping you with program evaluation and performance measurement activities? Please provide examples if possible.

From the WSC Recorded Training sessions, what was <u>least useful</u> in helping you with program evaluation and performance measurement activities? Please provide examples if possible.

What suggestions do you have for making the recorded sessions more useful for your program/organization?

How useful were each of the following WSC Training and Technical Assistance Components	**Not at all**	**A little**	**Somewhat**	**Very**
WebEx Site Visit with the WSC training staff	❑	❑	❑	❑
Recommendation Report from WSC after Site Visit	❑	❑	❑	❑
Phone/WebEx follow-up program solving session	❑	❑	❑	❑

What is one takeaway learning or action you took as a result of the WSC Training and Technical Assistance that helped or will help you with program evaluation in the future?

What suggestions do you have for making the site visit, program report or follow-up problem solving session more useful for your program/organization?

APPENDIX 4: Listening Session Guide

Year 2 Site Visit Listening Sessions: Structure for Taking Notes
January 2020

A. **Extent to which visit follows planned agenda**

B. **Topics covered/materials referenced**

C. **Level of engagement/participation by program** (share examples, evidence—eg questions asked, etc.)

D. **Examples of tough questions/sticky points of conversation and how resolved**

E. **At end of conversation, next steps identified?**

F. **Overall pluses of visit**

G. **Overall deltas of visit**

H. **Does this visit seem like could only be done by Robert/expert or could be replicated by WSC at broader scale?**

I. **Does it seem like information is presented in a way that the organization can think about and use evaluation more broadly or across a range of different projects or does it seem tailored only to the needs of WSC/Americorps?** (note examples where possible)

J. **Anything additional useful to share with evaluation team before interview with this site?**

APPENDIX 5: WSC Program Evaluation – Year 2 Treatment Participant Interview

My name is Patricia Keenan and I am working as part of the research team for the overall evaluation of the WSC Training and Technical Assistance intervention. As part of the evaluation we are interviewing individuals who participated in the four training webinars, the WebEx visit with Robert, review of a recommendations report, and follow-up call. In this interview we will ask about your impressions of the training and technical assistance, what you learned, and your plans for implementing some of the practices or recommendations discussed. The interview should take about 45 minutes. Please be as honest as possible and be assured we will not share any identifiable information with the WSC staff. Please let us know if you have any questions as we talk!

Would be OK if we record this interview to ensure accuracy of your responses? YES/NO

- How would you describe your role with the program/organization/agency?
- Now I'm going to ask you some questions about your impressions of the training and technical assistance. Do you think you gained some new knowledge or insights as a result of your participation? If yes, can you share an example or two? Are there areas where you feel your knowledge is the same?
- Have you experienced any changes in your confidence or ability to carry out different evaluation related activities – *probe for specific change – the four topics of the webinars were Program Alignment, Evidence of Effectiveness, Data Collection/Use, and Data Quality.*
- Did you or your organization implement any new or modified evaluation/data collection/data assessment practices as a result of participating in the training. (If yes), please tell me about what you did and who was involved. *[probe for ones that are specific to WSC requirements and if there are any that will help them beyond WSC]* Are there examples where your implementation of these practices has worked well? Are there examples where this implementation has not worked as well?
- Are you planning on implementing any additional practices in the future *(i.e., if they have not done these to date)*? *[probe for if these are specific to WSC requirements or go beyond that]*
- As a result of this training, do you think your organization will change the way you use data or evidence to inform decision making? To inform program design? If so, could you provide an example.
- In thinking about planning some new evaluation practices, what are some factors that could help with implementing them. What are some factors that you think could be problematic for implementation?
- Which components of the training *(i.e., webinars, WebEx meeting, recommendation report, follow-up call)* do you feel were the **most useful** to you/your agency *(probe on the specific different kinds of evaluation activities—in what ways? How?)*.
  Which components were **less useful** – anything specific about those components that were challenging, confusing, etc. *(probe to see if they felt there was too little/too much overlap between webinars and other training components)*
- Did you encounter any challenges or concerns over the course of the training intervention (*make sure to probe for specific areas of concern*). How did you overcome some of these challenges?
- Were there any areas or topics missing that you wish had been covered?
- What more do you wish you could learn about evaluation?

Thank you for your participation this interview!

APPENDIX 6: Staff Interview

WSC Staff Interview Guide Year 2
May 11, 2020

Thanks for joining the call today. The purpose of this interview over the course of the next hour is to learn from you and hear your perspective on implementing the intervention this year. This will help add context to the other data we are collecting and is similar to the interview we did with you last year.

I'll be asking you questions about 5 main areas: 1) your overall perceptions; 2) preparation; 3) elements of the intervention; 4) how this year compared to last year; and 5) thoughts around replicating and scaling in the future.

I'd like to record today's interview so that I can revisit our conversation for my own notes. May I have your permission to record our interview today? (If yes, I'll record video/audio via Zoom). Marc will be taking notes too and may jump in, if something comes up that he wants to hear more about.

Any questions? Ready to start?

Overall perceptions

1. Overall, how have you felt the intervention this year has gone? (eg overall successful, a mixed bag, overall unsuccessful).
    a. Are there particular elements (webinars, site visit, follow up reports, follow up calls) that you think were more successful or less? Why or why not?

2. Did some programs seem more engaged or to learn more than others, from what you've observed so far? Did you notice any patterns?

3. Were there sites that really didn't need this or did you feel like all of them could benefit in some way?

Preparation of site visits and rec reports

4. Tell me a bit about your preparation for the site visits
    a. How long did it typically take you to prepare?
        i. (Did prep for some kinds of programs take longer than others, based on their focus area (here I mean the AmeriCorps focus areas--eg environmental programs tend to have less evidence available to them than education ones) or on how complex their program is? How old it is or how well you know them? Some other factors?
    b. What about preparing the recommendation reports? That seemed to take a bit of time, and it seemed like you had less support than you'd originally planned.

Implementing Different Elements of the Intervention and This Year Compared to Last Year

5. Overall, how did you feel about implementing the intervention this year compared to last year? What were some of the main differences overall or for any of the specific components?

6. What are some of your overall takeaways from implementing site visits virtually this year? What went well? What was challenging?

7. From what you could tell, were there particular webinar topics or modules that seemed more or less well-understood by the programs?

8. Were there particular types of programs or staff that seemed to engage or understand the material more or less? Any hunches on why (e.g., their tenure with WSC, AmeriCorps, etc.)

Replication and Scaling

9. In your opinion, what are some of the most or least important elements of the intervention if you were to do this again in future?

10. Would you recommend offering the webinars in future or would you cover that content differently?

11. If you were to give advice to someone about to replicate your intervention in future, what would you tell them?

Final Takeaway Reflections

12. Anything else you'd like us to know or think about?

APPENDIX 7: Individual Knowledge Questions – Comparison: Treatment vs. Control

Evidence of Effectiveness

TABLE A1: Q1-When we speak of the evidence of effectiveness of an intervention we refer to *(choose one answer)*

	CONTROL		TREATMENT	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
Studies indicating that a program is likely to lead to a specific effect or outcome	55.3%	59.6%	63.4%	80.5%
Studies indicating that program staff believe that the intervention is working	8.5%	4.3%	4.9%	2.4%
Studies indicating that a program carried out the intervention as intended	31.9%	29.8%	22.0%	17.1%
Studies indicating that a program met the expectations of the funding agency	4.3%	6.4%	9.8%	0.0%

TABLE A2: Q2- When building your program's evidence of effectiveness …. (*choose one answer*)

	CONTROL		TREATMENT	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
past performance measurement data on outcomes cannot be used as evidence	6.5%	2.2%	4.8%	4.8%
demonstrating positive change on an outcome pre to post may count as preliminary evidence	67.7%	73.9%	69.0%	78.6%
a program's life cycle determines the standard for preliminary evidence	8.7%	6.5%	23.8%	14.3%
results from studies of other programs should not be considered	17.4%	17.4%	2.4%	2.4%

TABLE A3: Q3- Which of the following best describes higher levels of evidence on the continuum of Evidence of Effectiveness *(choose one answer)*

	CONTROL		TREATMENT	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
Research studies at higher levels of evidence have larger sample sizes	6.7%	2.2%	7.1%	0.0%
Research studies at higher levels of evidence look at behavioral rather than knowledge or attitudinal outcomes	24.4%	28.9%	33.3%	11.9%
Research studies at higher levels of evidence include a comparison or control group	66.7%	66.7%	54.8%	83.3%
Research studies at higher levels of evidence do not include qualitative data.	2.2%	2.2%	4.8%	4.8%

TABLE A4: Q4 -Which of the following are the best examples of a study that will meet the standard of preliminary evidence for an outdoor science education program for middle schoolers (*choose all that apply*)

	CONTROL		TREATMENT	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
A study on how the program was implemented with 85 middle schoolers	22.9%	10.4%	9.5%	14.3%
A study demonstrating that 85 program middle schoolers enjoyed the program	8.3%	8.3%	0.0%	4.8%

A study demonstrating that 85 program middle schoolers knew more science concepts than a control group of 85 middle schoolers in the spring	39.6%	64.6%	54.8%	59.5%
A study demonstrating increased knowledge among 85 program middle schoolers from fall to spring	79.2%	70.8%	85.7%	76.2%
% Totally Correct	***22.9%***	***35.4%***	***40.5%***	***33.3%***
Average Number Correct (out of Max of 4)	***2.88***	***3.17***	***3.30***	***3.17***

Program Alignment

TABLE A5: Q1- You are delivering a parenting workshop to a group of participants over a three-month period. Look at each statement and choose the part of a logic model where it would best fit:

	CONTROL		**TREATMENT**	
Correct answer in bracket	**Baseline**	**Follow-Up**	**Baseline**	**Follow-Up**
40 parents complete eight weekly sessions [OUTPUT]	36.2%	48.9%	38.1%	50.0%
30 parents report increased knowledge of support services available to their family [OUTCOME]	53.2%	66.0%	45.2%	59.5%
The staff present information on positive discipline techniques [ACTIVITY]	67.4%	63.0%	61.9%	66.7%
3 full time staff manage and deliver the program [INPUT]	76.6%	63.8%	70.7%	75.6%
50 teen parents increase their use of positive discipline techniques [OUTCOME]	93.6%	87.2%	95.2%	85.7%
% Totally Correct	***3.25***	***3.30***	***3.09***	***3.38***
Average Number Correct (out of Max of 4)	***29.8%***	***36.2%***	***21.4%***	***38.1%***

TABLE A6: Q2-Which of the following is the best example of a thorough intervention statement? *(choose one answer)*

	CONTROL		**TREATMENT**	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
Our program distributes informational materials to interested community members	0.0%	0.0%	0.0%	0.0%
Over 200 of the students in the 5 different after school tutoring programs demonstrate increased academic performance over the school year	44.7%	25.5%	34.1%	29.3%
The program provides substance abuse prevention training to 300 middle school youth through 8 2-hour weekly sessions in 5 sites	55.3%	72.3%	65.9%	70.7%
Our program gives out weekly vouchers to homeless families	0.0%	2.1%	0.0%	0.0%

TABLE A7: Q3-A program provides financial planning workshops to low income families. Which statements below represent outcomes that align with the program activity (*choose all that apply*)

	CONTROL		TREATMENT	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
80 participants increase their knowledge of how to set up a savings account	93.8%	85.4%	88.1%	83.3%
70 percent of participants who enroll are from low-income households	22.9%	14.6%	9.5%	11.9%
100 individuals attend workshops in three different locations	10.4%	10.4%	11.9%	4.8%
200 participants demonstrate the skills needed to balance personal budgets	89.6%	85.4%	88.1%	92.9%
200 participants demonstrate improved attitudes toward parenting	20.8%	22.9%	11.9%	23.8%
5 staff train 3 AmeriCorps members in how to deliver the program	2.1%	2.1%	0.0%	2.4%
Average Number Correct (out of Max of 6)	***5.27***	***5.21***	***5.42***	***5.33***
% Totally Correct	***47.9%***	***56.3%***	***57.1%***	***47.6%***

TABLE A8: Q4-You are assessing the quality of the logic model above. Which of the following statements are correct? *(choose all that apply)*

	CONTROL		TREATMENT	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
"100 adults complete job search training" SHOULD NOT be an input	91.7%	75.0%	90.5%	88.1%
"100 adults complete job search training" SHOULD NOT be an output	10.4%	18.8%	7.1%	9.5%
"80 adults demonstrate new job search skills" SHOULD NOT be an output	60.4%	60.4%	47.1%	59.5%
"150 adults increase their self-confidence in their job search skills" is NOT APPROPRIATE because self-confidence is not a possible outcome.	16.7%	20.8%	7.1%	7.1%
"150 adults increase their self-confidence in their job search skills" is NOT APPROPRIATE because a change for 150 is not a possible outcome	64.6%	58.3	69.0%	71.4%
The outcomes DO NOT ALIGN with the program design	14.6%	14.6%	16.7%	19.0%
Not all of the short-term outcomes are short term	85.4%	79.2%	95.2%	92.9%
A 5-year outcome should not be in a logic model	14.6%	8.3%	19.0%	16.7%
Average Number Correct (out of Max of 8)	***6.46***	***6.10***	***6.52***	***6.60***
% Totally Correct	***18.8%***	***14.6%***	***21.4%***	***26.2%***

Data Collection and Use

TABLE A9: Q1-Which of the following is NOT true about the process of performance measurement? *(choose one answer)*

	CONTROL		TREATMENT	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
It involves the collection of output and outcome data	4.2%	2.1%	2.5%	0.0%
It measures performance against targets	6.3%	0.0%	5.0%	5.0%
It gathers data that is different from the data a program uses to inform decision making	60.4%	79.2%	72.5%	82.5%

It allows the user to test personal observations with empirical data	29.2%	18.8%	20.0%	12.5%

TABLE A10: Q2-It is important to use data to inform decision making about the program. If your results show that you do not meet your outcome targets, some appropriate next steps could be *(choose all that apply)*

	CONTROL		TREATMENT	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
Increase program outreach to other target populations	64.6%	60.4%	61.9%	47.6%
Expand the same services to more participants	47.9%	45.8%	35.7%	33.3%
Consider finding another evidence-based practice to try and replicate	83.3%	66.7%	90.5%	81.0%
Find a new way to deliver similar services	85.4%	81.3%	88.1%	78.6%
Consider ending the program services	33.3%	27.6%	26.2%	40.3%
Completely adjust your logic model to fit your results	6.3%	6.3%	0.0%	
% Totally Correct	***18.8%***	***12.5%***	***19.0%***	***11.9%***
Average Number Correct (out of Max of 6)	***4.17***	***4.08***	***4.55***	***4.38***

TABLE A11: Q3-Your program has an outcome that participants in an Earth Day event increase their use of environmentally practices after the event. Which of the following below would be the best way to collect data to inform you about that outcome? *(choose one answer)*

	CONTROL		TREATMENT	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
Observation of the actions of individuals at a future event	4.2%	4.2%	4.8%	2.4%
A focus group three months after the event asking individuals what they liked about the event	2.1%	4.2%	0.0%	0.0%
A self-report survey completed 3 months after the event asking them what they know about different practices	16.7%	22.9%	19.0%	14.3%
An interview 2 months after the event asking them about their daily conservation practices	77.1%	68.8%	76.2%	83.3%

TABLE A12: Q4-A change in how someone felt about something would be an example of an *(choose one answer)*:

	CONTROL		TREATMENT	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
Conditional outcome	2.1%	6.3%	2.4%	4.9%
Knowledge output	2.1%	2.1%	0.0%	0.0%
Attitudinal outcome	81.3%	85.4%	90.2%	92.7%
Behavioral demonstration	10.4%	4.2%	4.9%	2.4%
Attitudinal input	4.2%	2.1%	2.4%	0.0%

TABLE A13: Q5-Thinking about the <u>previous question</u>, which of the following would be appropriate ways of measuring whether that change occurred *(choose all that apply)*

	CONTROL		TREATMENT	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
Pre/Post Knowledge Assessment	22.9%	19.6%	21.4%	11.9%
Observation of the individual's actions	39.6%	33.3%	26.2%	31.0%
Open-ended interview with individual	66.4%	68.8%	64.3%	64.3%
Pre/Post Self-report survey	91.7%	91.7%	83.3%	93.8%
Review of a case manager's case notes about things worked on with the individual	22.9%	27.1%	4.8%	16.7%

% Totally Correct	***20.8%***	***20.8%***	***26.2%***	***21.4%***
Average Number Correct (out of Max of 5)	***3.67***	***3.85***	***3.95***	***3.78***

<u>High Quality Data</u>

TABLE A14: Q1-Your program offers environmental education workshops for high school youth helping them understand stewardship practices to save water and energy. You are interested in determining whether their knowledge of these practices increases over time. Which of the following would be a VALID method for measuring this outcome: *(choose all that apply)*

	CONTROL		**TREATMENT**	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
One survey completed at the end of the workshop asking them their current knowledge of the practices	4.2%	2.1%	7.1%	2.4%
A pre-post survey testing knowledge of these practices before and after the workshop	95.8%	100.0%	100.0%	95.2%
Site visits to observe youth trying the practices before and after the workshop	77.1%	47.9%	66.7%	64.3%
An interview with selected participants at the end of the workshop asking about specific elements of the practices they know more about after the workshop	60.4%	52.1%	69.0%	45.2%
A tracking sheet indicating whether the youth attended a specific workshop	12.5%	14.6%	11.9%	7.1%
% Totally Correct	***6.3%***	***18.8%***	***11.9%***	***14.3%***
Average Number Correct (out of Max of 5)	***3.62***	***3.87***	***3.83***	***3.67***

TABLE A15: Q2-Which of the following would be a concern with the <u>consistency</u> of the data *(choose one answer)*?

	CONTROL		**TREATMENT**	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
The program uses one survey to measure program outputs and a second survey to measure program outcomes	4.2%	10.4%	4.8%	2.4%
The program only collects data from one of the four sites involved in the intervention	22.9%	8.5%	7.1%	7.1%
The staff use an observational tool to measure changes in client knowledge	4.2%	6.3%	0.0%	0.0%
Staff at one location use one survey to measure outcomes; staff at second location use a different survey to measure the same outcomes	68.8%	75.0%	90.5%	90.5%

TABLE A16: Q3-When reviewing some performance measure results, we find that the number of individuals who achieved an outcome well exceeded the number of individuals participating in the program. In this case, we are most concerned with: *(choose one answer)*

	CONTROL		**TREATMENT**	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
The validity of the results	38.3%	46.8%	23.8%	26.2%
The accuracy of data collection and analysis	55.3%	51.1%	69.0%	66.7%
The consistency of the data	2.1%	2.1%	0.0%	4.8%
The generalizability of the results	4.3%	0.0%	7.1%	2.4%

TABLE A17: Q4-Which of the following measures would inform you of whether your program has implemented program services as intended *(choose all that apply)*

	CONTROL		TREATMENT	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**
The number of clients who would refer their friends to your program	33.3%	29.2%	35.7%	26.2%
The dosage of services provided to each client	33.3%	45.8%	35.2%	42.9%
Each client's level of gain in knowledge	87.5%	77.1%	85.7%	88.1%
A staff member's willingness to provide more services for the agency	4.2%	8.3%	4.8%	11.9%
The frequency of delivery of the program activities	31.3%	52.1%	38.1%	45.2%
% Totally Correct	***8.3%***	***10.4%***	***11.9%***	***7.1%***
Average Number Correct (out of Max of 5)	***2.40***	***2.83***	***2.47***	***2.62***

APPENDIX 8: Follow-Up Survey Feedback Data – Assessment of Recorded Session Components

Most Useful

I'm not sure exactly what is meant by the WSC Recorded Training sessions, but talking one on one with Robert was helpful to get a better sense of what data is most important and useful for WSC, and what is truly required and what is optional for us to collect.
Each training was very beneficial, but I especially learned a lot through the training checklists and skill building exercises we were able to do following each training session.
We worked with our WSC staff to refine some of our data-collection tools and we reduced unnecessary data collection. He shared some models for collecting commitment data that will be helpful for us to redesign this tool.
I thought most of the examples were helpful, especially the examples on how to write a logic model. I had very little understanding of how to use a logic model but now I feel good about it.
It was helpful to see more detailed explanations and examples of outcomes and outputs. WSC has changed criteria, or at least communication around criteria, frequently and it has made it challenging.
I can't remember the recorded sessions
I barely remember these trainings as they were several months ago. But I do recall some graphics being used to help explain each step and it seemed to make sense at the time.
All of the trainings were good overall. The definitions of the terms, visuals, and charts were extremely helpful.
Being new to this position and managing and AmeriCorps Service Member, it was very useful to be sent all the necessary documents and research my predecessor developed prior.
defining terms was most helpful. When we next have an opportunity to reevaluate our logic model we will be better equipped to refine and revise in a way that aligns with WSC.
Detailed conversation with Robert Brader about our specific program and materials we had submitted in support of our AmeriCorps application
Since I am pretty new to this side of WSC, it was really beneficial for me to get a better idea of the vocabulary and idea behind the performance eval/management process.
It was just a good overview of what we should be doing and it was helpful to be asked to look over it again.
Performance measures
Discussion with Robert Brader helped apply training sessions to our specific situation.
Having all of our materials reflected back to us in an organized and concise manner with input as to their accuracy and effectiveness
Hearing tangible actions we could take for improving.
Introduction to various relevant research on tools known to be effective with other similar programs.
Good to have transcript and video. Liked checklists. Enjoyed that I could complete this at my own pace.
examples were helpful
Examples of discussion topics
A good refresher on logic models & data collection.
It was helpful to look at types of assessments that we do currently and see how that fell on the continuum of the data being viewed as reliable.
It helped us understand the framework that WSC applies and how they want it articulated.
I think it is important to measure the effectiveness of our programs to provide changes/feedback to our funders. It is also important to measure over-time so we can better develop our programs.
Difference between outputs and outcomes. Learning about the types of problems with data. Follow up knowledge tests. It seems like a fair middle ground to provide information without arranging the time and travel for in person training.
I'm sorry but I am completely out of time to continue to answer these questions when there are much more pressing needs within my community!
Seeing examples
Specific examples; applying the material.

Least Useful

Again, I'm not sure what is meant by WSC Recorded Training sessions, but if it is referring to the online trainings, it was somewhat repetitive, time-intensive, not very engaging, and was full of jargon.
All training sessions were beneficial
Establishing a baseline of evaluation language prior to the recordings would have been helpful. I found myself pausing the recordings and looking up things separately because there was a lot of information and terminology.
I don't recall anything that wasn't useful at the time, but then again, I couldn't tell you anything that happened in the trainings other than the colors red and blue were used to make rectangular boxes with words in them.
I think they were all helpful. It would be good for AmeriCorps to remember that a lot of staff watching and utilizing the info from the trainings are new to many of the terms (not necessarily the actions) in program evaluation. At times there may be confusion in conversations and communications as staff learn the terms
I would say it was all very useful for me because as stated above, this was all fairly new for me.
information regarding the characteristics of higher and the highest level research. During the recorded session I was left with the impression that WSC may be asking us in the future to conduct research at a level that is beyond our foreseeable capabilities. While WSC staff assured us that this was not the case, I was left wondering why they felt it was relevant to our program evaluation and performance measurement activities that we were being trained on if there was no expectation now or in the future that we implement this knowledge.
Is this question asking about the conversation with Robert Brader or the recorded data collection trainings on the WSC site? My answers regard the conversation with Robert Brader. Some examples didn't feel relevant to our site. The conversation was not always concise.
It has been a while since I watched these training sessions, so I don't recall which parts weren't helpful.
Just finding the time to go through them.
Logic Models. I've had a lot of trainings and practice with this already.
Lots of info focused on tracking changes in knowledge, attitudes, behaviors, etc., which is not data that our program tracks.
Repetitive information.
Reviewing our current logic model & talking about new ideas that we had. It was able to get us going on thinking of new ways to collect data for what our members are doing & how we could change what we collect/our goals to better align with the new needs of our clients.
Some of the information was common knowledge regarding study design.
There was a lot of review of the pre-survey material. That portion could be condensed. The time could have been better spent on more applications.
While we are accountable to our funders and required to share photos, reports of our work and measure our progress - each funder has a unique set of criteria for measuring our effectiveness. I do respect the required training by WSC but in this circumstance it did not seem helpful to participate in this training.

Suggestions

(Again referencing conversation with Robert). The fact that the conversation was personalized for and directly relevant to our program was quite helpful. I would encourage this to be done more frequently for organizations with AmeriCorps members.
Cannot think of anything. It was pretty customized already.
Given the program calendar I think it would be most helpful to share these resources during the spring and summer when organizations are working to improve their data collection strategies for the coming year.
I primarily used the slide decks, as those fit me better. Sometimes the audio in trainings isn't always the best
I would have appreciated these trainings earlier on in my onboarding process to inform my approach to the issues my organization has been navigating

I would have liked to have a little more time to dedicate to these surveys and check-ins, but they came at a pretty busy time for us. Summer would be better.
In a situation where particular sites are selected for training related to their program evaluation and performance measurement activities it would be most helpful for staff administering the training to have observed the activities as carried out by the folks they are training. If recorded sessions are the preferred method used to administer these trainings, then observations of the activities carried out by those being trained could inform those recorded sessions.
in person meetings for those of us in the area would be great!
It would be helpful to discuss some of the challenges in collecting and analyzing data. When participants are difficult to reach or making sure your collection methods are culturally relevant. When funders want you to track different things or requirements change.
Make it simpler, shorter, and more engaging.
More information tailored to at-risk ecosystems or other performance measures that don't rely on tracking changes in knowledge, attitudes, behaviors, etc.
More opportunities for interaction with the material.
Sometimes the volume/sound quality on the recorded videos is very bad and it's distracting/difficult to watch.
These would have been more useful when we were submitting our original RFA to ensure that we were as accurately and completely as possible.
They all said that there was a quiz at the end, yet none of them had a quiz. Perhaps a quiz would have helped me remember the content of the trainings.
They could be more concise. Seemed like a lot of fairly obvious material.
This process has been extremely difficult and time consuming in a way that feels unnecessary and not super helpful. Every program requesting an AmeriCorps is understaffed and extremely busy, it is why they have asked for an AmeriCorps. This time intensive evaluation has been a burden for us and we will not be participating in the program last year partly due to the reporting and time intensiveness of participating in the evaluation.

APPENDIX 9: Follow-up Survey Feedback Data – Assessment of Intervention Components

Key Takeaways

Better pre/post evaluations and the importance of logic models.
Better understanding of the breakdown of data categories and continuum of evidence of effectiveness.
Clearer understanding of WSC's reporting priorities, which will help us better focus our own reporting priorities and eliminate unnecessary data collection.
Focus on qualitative data
Have more/better research studies regarding the types of program delivery and its effectiveness
Helped me shorten the logic model which was very helpful
How to beef up our quarterly reports, and we brainstormed some really helpful ideas to better tailor our outputs and outcomes with member activities.
I appreciated learning about how other organizations that are completing similar projects are assessing their programs. We often reinvent the wheel in environmental education, and the knowledge of what other groups are doing will help us with program evaluation in the future.
I think it is necessary to reevaluate our plastic pledge as well as how we document our data using acreage. It makes more sense to track how many pounds of marine debris we collect during our cleanups.
I'm sorry I don't recall much. But I do appreciate the need to inform us when the annual overhauls to the program occur.
In the short term the WSC training helped me to understand mistakes I was making in quarterly reporting that will be corrected moving forward
Knowing that we can work together with Robert to come up with performance measurements that fit our program and WSC's needs.
More data and references in logic model/problem statement.
Nothing particularly new, as we had been working on the recommend items already. Recommendation to better align our logic model with program description is a good suggestion.
Our recommendation report was very helpful! The biggest takeaway, by far, was the understanding that WSC is open to conversation about data collection strategies. For years there has been a perceived desire that AmeriCorps is seeking quantity over quality with its data. This process really highlighted for us that there is a strong preference for quality and a willingness from WSC to discuss the best way for our program to report on our work. We're excited to focus our interventions and data collection strategy on long-term impact in the coming program year.
Our WSC rep provided some new studies that we were not familiar with that have bearing on our kind of interventions.
Rethink how we assess change in behavior.
Starting the process on revisiting our program goals & interventions for our AmeriCorps members that align more with the needs of our clients now. This would also change what we collect & report on to WSC.
Suggested edits to the logic model like adding dosage information.
That we need to periodically reevaluate our logic model, program delivery, and data collection to make sure we are getting the inputs, outputs, and outcomes as focused as possible.
To choose a project that more closely aligns with what we are currently doing. Our final follow-up call with Robert was very positive in that we suggested if we were to apply for a future program with the WSC, we would add Military families to the focus, not just veterans which our current program is focused on.
To-dos from recommendation report
Transitioning from knowledge-based assessments to behavior based assessments. We still plan to use some knowledge-based assessments to supplement our observational assessments when reviewing the individual portions of the program to aid in fine tuning classes. These assessments will supplement the behavior based pre and post surveys

Very helpful to think through data collection methodology.
Ways to refine our logic model and increase accuracy in data collection
We are going to review the recommendations and use them to make changes to our logic model next time we need to re-submit our WSC application (next year).
We decided to discontinue the use the of the UCLA Loneliness Scale. I had felt that it was harmful to those we serve.
We have changed the data we are collecting and have tried a few of the suggestions from the Webex meeting.
We need to identify what tools we want to use for the next program year
We will be looking at building out a more robust logic model for the entire program.
We will figure out a way to send a follow-up to parents and teachers to extend the lessons into their daily lives and connect to the natural settings they encounter in their everyday activities.

Suggestions

Be more prepared
Can't think of anything. We covered a lot in the recorded session that I took notes on and thus receiving the report reiterated those points, but the attached studies were very interesting.
For me it was very useful and will help me for our next term of AmeriCorps. I believe I will have more suggestions and input as I become more comfortable with the program. Thank you!
Had trouble logging in at first, but afterwards went well.
I believe that this information is valuable, however it was time consuming and we didn't budget for two staff to spend the hours on this. It would be helpful to know early on when organizations are awarded the placement of a member that they were selected for this process so that this time could be budgeted for.
I had a really hard time with some of the wording used. Perhaps a bit more time converting the institutional language to plain English might have made the concepts stick better.
I think having the ability to conduct our meetings over the phone was incredibly helpful - we don't have funding to travel or attend meetings in person in Olympia and we were grateful for this option. Thank you!
I think it would have been easier to participate in the summer. The timing of the study is a real crunch time at the organization, and I wish I'd had more bandwidth to let the lessons sink in and reflect. I feel like I had to rush.
I thought that having examples of data helped illustrate. I also liked having some time to review the report before the follow-up call.
If the site visit was in person and WSC staff observed the activities at the site I believe it would deepen the learning for site staff.
I'm really unsure how to improve. I understand it's logistically challenging to "site visit" so many sites at great distances, but we found it a challenge to communicate with our WSC person over the phone. In person would have been more effective for us.
It felt way over-formalized and excessively time-consuming. Having one-on-one coaching assistance available as needed would be much more helpful and appreciated than this entire elaborate process being required irrespective of need.
It was helpful. It would be great to have this type of conversation every couple of years with WSC.
Maybe conduct earlier in the program year?
needed meeting options further out- our ED's schedule is packed, fitting in a meeting with only 1-2 weeks notice is challenging. Goal for webex wasn't very clear beforehand. Having an agenda and questions emailed prior would have been helpful.
Sample surveys are extremely helpful, having a number to review is helpful.
Schedule all meetings closer together a I felt there was too much time in between each. THANK YOU Robert for all your time and assistance with this process!!! :)
Shorten the timeline. The time between all the elements was so long, it was hard to remember/keep track of what we talked about.

The recorded training could be designed to be more interactive so that it responds to different sites' needs. For instance, an initial question that asks what parts of the presurvey material they want to focus on, and what issues the site finds most challenging might help streamline the recorded sessions so that they don't feel like excessive review and are more geared towards sites' different levels of understanding and needs. The WebEx session would have been more useful and could have been more streamlined had sites been asked what they considered to be their main challenges, and what they wanted to focus on during the meeting. Receiving some kind of agenda prior to the session would have been very helpful in terms of knowing what to expect.
The trainings ahead of time were unnecessary and not helpful to lead that discussion.
Understand what our members and organization do before doing a program report.
We received the recommendation report a long time after the initial site visit and then had our follow-up call shortly after receiving the report. It would have been more beneficial to get the report quickly following the visit and then have a longer period to review before the follow-up call. Additionally, it would be very helpful to have an online resource library of studies, preliminary evidence, etc. available to all sites.