

# CAPACITY ASSESSMENT



U.S. DEPARTMENT *of* STATE

April 2022





# U.S. DEPARTMENT *of* STATE

## TABLE *of* CONTENTS

EXECUTIVE SUMMARY.....5

STATE DEPARTMENT POLICY AND TOOLS .....7

    Capacity Assessment Development.....7

    Stakeholder Engagement.....7

FINDINGS .....7

    Maturity Model Ratings .....8

    Coverage.....9

        Staffing.....9

        Activities and Systems.....10

        Evaluation Coverage .....11

        Performance Monitoring and Program Design .....12

        Research and Analysis Coverage .....14

    Methods .....15

    Quality and Effectiveness.....18

        Disseminating Good Practices and Findings .....21

        Day-to-Day Operations and Learning Needs .....22

        Capacity-Building Activities .....23

    Independence .....24

CONCLUSIONS.....25

    Agency Capacity.....25

ISSUES FOR CONSIDERATION.....26

APPENDIX A: METHODOLOGY.....29

    Maturity Model.....29

    Desk Review of Existing Data.....29

    Process .....29



Survey Design.....30

    Conceptual Map and Measurement Priorities .....30

    Process .....30

Procedures .....31

    Data Collection .....31

    Data Analysis.....32

    Limitations and Mitigation .....32

APPENDIX B: FY 2023 ANNUAL EVALUATION PLAN .....33

APPENDIX C: STAKEHOLDER ENGAGEMENT .....36

APPENDIX D: MATURITY MODEL.....37

APPENDIX E: CAPACITY ASSESSMENT SURVEY .....39





U.S. DEPARTMENT *of* STATE

Capacity Assessment for Statistics, Research,  
Evaluation, Performance Monitoring, and Other  
Analysis



## EXECUTIVE SUMMARY

This assessment reviews the Department of State's capacity to generate and apply evidence through performance monitoring, evaluation, statistics, and research and analysis. This baseline will help the Department measure its improvement in these areas and contribute data to inform a tailored capacity-building plan for the Department.

This report analyzes capacity assessment data collected in fulfillment of the Foundations for Evidence Based Policymaking Act of 2018 (Evidence Act) by the Office of Foreign Assistance and the Bureau of Budget and Planning. The Department's co-Evaluation Officers directed this work, and it was carried out by subject matter experts from both offices.

Key findings about the Department of State's capacity to generate and apply evidence included:

- **Performance Monitoring and Evaluation Growth:** Roughly 88% of Department bureaus and independent offices collect performance monitoring or project indicator data. They also observed growth in evaluation activity over the last few years. However, many of these bureaus collect data on some but not all of their project indicators and performance goals. Five of 44 (11%) bureaus and independent offices may need technical assistance to fully comply with the Department's policy on performance monitoring and design.
- **Evidence-Building Staff Coverage:** There are roughly 420 Department staff dedicated full-time to evidence building, but 11 of 44 (25%) bureaus or independent offices reported they have no full-time evidence building staff. Of the other 33, 19 (58%) report that they have an insufficient number of staff to conduct evidence building.
- **Evaluation Coverage:** 72% of bureaus and independent offices have commissioned external evaluations in the past three years, as reported through survey data and evaluation databases.
- **Statistics:** The Department's maturity is at the evolving level in this domain, recognizing the Department of State is not a designated statistical agency. Respondents to this assessment generally look at statistics as an analytical tool to apply to their bureaus' own data, rather than a distinct set of programs. Further analysis of survey data reveals that fundamental techniques are present in most units, like fundamental statistics (65%), but that more nuanced skillsets, like Bayesian techniques (28%), were less likely to be observed.
- **Independence:** 60% of Department bureaus and independent offices indicated that they agree or strongly agree that their bureau mitigates inappropriate influence, so that evidence from activities is systematically and fairly considered regardless of the findings.
- **Research and Analysis Effectiveness:** 17 of 44 (39%) of bureaus and independent offices rated their research and analysis in the two lowest levels of maturity (not performing or evolving).





- **Sharing Evidence Findings:** A majority of bureaus and independent offices share evidence findings within their bureau (80%) or within the Department (65%). A slight minority of bureaus (42%) reported often disseminating their evidence to other federal agencies and to Congress (35%).

Based on the key findings in this assessment, the team recommends:

- Bolstering the evidence-building culture in the Department through targeted hiring, including specialists in data analysis, evaluation, research, and learning and by ensuring that performance is measured against capacity to integrate data and learning into strategic planning.
- Strengthening evaluation in Department-wide processes such as resource strategic reviews, Joint Strategic Plan (JSP), among others.
- Promoting data literacy across the Department and connect bureaus with the Center for Analytics and other data analytic tools in the Department.
- Integrating other forms of evidence within registries and invest in analyses of evaluations with the goal of building information and data that a wider audience across the Department can access.
- Reinvigorating professional development and training opportunities for research, evaluation, and learning staff within the Department to ensure familiarity with data and analytics skills.
- Sharing evidence building activities across bureaus, including data, analysis, and evaluations collected in some bureaus with relevance or adaptative potential to other bureaus.
- Recognizing evidence-building funding requirements across the Department through budget formulation and operational plans as well as through senior leader prioritization.



## STATE DEPARTMENT POLICY AND TOOLS

In 2012, the Department introduced its first evaluation policy, and in subsequent years expanded it to include program design, performance monitoring, and changes based on the Foreign Aid Transparency and Accountability Act of 2016. This expansion codified and strengthened existing practices for program and project management, monitoring, evaluation, and using data for learning.

To support the implementation of the policy, the Department provided training courses for staff, including strategic planning, performance management, managing evaluations, and evaluation design. In addition, the Department created communities of practice for evaluation, program design, and performance management specialists. The Department offered technical assistance to bureaus and independent offices on how to develop theories of change, indicators, evaluation scopes of work, and strategic plans. It also encouraged and supported adoption of best practices through a contract that facilitated access to independent evaluators.

### Capacity Assessment Development

#### Stakeholder Engagement

Engaging stakeholders has been a staple of the capacity assessment process. From the earliest stage, the team mapped stakeholder groups, planned for different types of engagement, and used that information to disseminate information and seek feedback on its efforts.

The team has engaged the Department's evidence-building community, Department leadership, and external stakeholders including think tanks, implementing partners and the Office of Management and Budget (OMB). A launch event for the entire evidence-building community was held on July 30, 2020, to brief them on the Evidence Act, its benefits, and implications for the Department. After that, a working group was formed of more than 40 subject matter experts from across the Department to help shape the data collection tools and plans for data collection. A series of working group webinars were held to explain, review, and gather input from working group members on the drafts of the maturity model, an organizational capacity measurement tool, in addition to input on the plan for data collection. The team also engaged the Chief Data Officer/Acting Statistical Official, who is responsible for data governance and lifecycle management across the Department, to review the survey instrument and data collected to inform the Enterprise Data Strategy (EDS). Throughout this process, the team has engaged bureaus' leadership on research plans, tool development, and communication strategy. The Department's Evaluation Officers have also shared progress updates with executive branch and legislative branch stakeholders.

## FINDINGS

Analysis and findings for this report are introduced first with an overview of the maturity model summary ratings and then by Evidence Act criteria including coverage, methods, quality, effectiveness, and independence. These findings include data compiled from existing sources and a Department-wide survey administered to internal evidence-building experts.

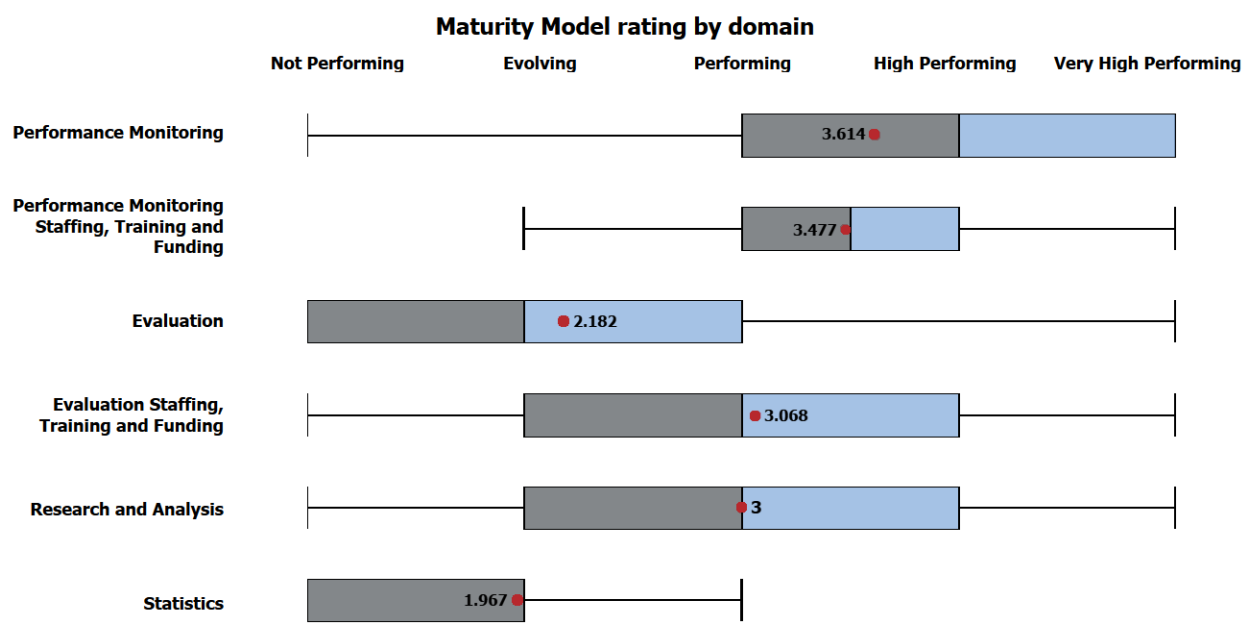


## Maturity Model Ratings

The maturity model summary ratings in Figure 1 show the Department as performing in the higher range of the ‘Performing’ rating for performance monitoring and the staffing, training, and resources to support it and ‘Performing’ or ‘Evolving’ for evaluation, statistics and research and analysis.

These ratings demonstrate that the Department has focused on the program design necessary for generating plans to collect monitoring data and invested in the policy, training, staffing, and support to continue this work. As further analysis will also show, it also highlights that there are numerous parts of the Department that focus on this kind of evidence-building as it is the core of their daily work. The evaluation domain summary ratings demonstrate that the Department needs to make progress in producing evaluations and, on average, is staffed and resourced to do so. However, while this provides a valuable snapshot of average bureau-level capacity at the Department, important progress remains to be made in staffing depth, quality of evaluation work, and encouraging utilization of evaluations to improve agency performance.

Figure 1. Maturity Model ratings



This above graph, known as ‘box and whisker’, shows a 5-number summary of data. The **minimum**, which is the lowest datapoint in the dataset, is the leftmost point on the chart on the end of the left “whisker”. The **First Quartile**, which represents the 25% mark of the data, is the left edge of the “box”). The **median**, which represents the 50% mark or center of the data, is the line between the left and right sides of the “box.” In the charts shown above, the median lies where the gray and blue sections of the “box” meet. The **Third Quartile**, which represents 75% mark of the data, is the right edge of the “box”). The **maximum**, which is the highest datapoint in the dataset, is the rightmost point on the chart on the end of the right “whisker.”



## Coverage

### Staffing

To list activities and understand coverage (i.e., where evidence-building activities are happening and who is responsible), we sought to understand how many bureaus and independent offices have divisions or offices that are dedicated to or have core functions in evidence-building (i.e., those defined in job descriptions and formal responsibilities, not program officers for whom routine performance monitoring is part of their project oversight duties). Based on this question, respondents reported an average of 2.1 divisions or offices (median: 1.5). In total, 96 evidence-building divisions and offices were reported across the Department, with numbers per bureau ranging from zero to eight. Two bureaus reported that they did not have an office or division dedicated to evidence-building. While most bureaus or independent offices would likely benefit from an evidence-building unit, the Department would need to account for the varying sizes and scopes of these organizational units to understand these implications.

As activities and infrastructures differ across bureaus and independent offices due to their varying sizes and scopes, the team took the opportunity to understand which actors are involved in evidence-building work across the Department. Nearly all respondents (98%) noted that federal employees are involved in evidence-building activities in the bureau or independent office. Additionally, 86% of bureaus and independent offices employ onsite contractors responsible for evidence-building work. A majority (66%) also noted that implementing partners carry out evidence-building activities, which are tied to conducting performance monitoring activities for the Department's contracts and grants, especially those connected to foreign assistance funding. A slight minority (39%) of respondents noted that their units work with academic partners to generate evidence.

In addition to understanding the number of evidence-building units carrying out activities across the Department, the team enumerated those staff that have core responsibilities for this type of work. We asked respondents to specify staffing numbers based on the level of effort spent on evidence-building activities. Table 3 highlights the number of staff members—regardless of hiring mechanism (e.g., onsite contractor, direct hire)—identified as having core responsibilities related to performance monitoring, evaluation, statistics, and/or research and analysis. Respondents received instructions to not include program officers who conduct routine performance monitoring as part of their project oversight duties as the intent was to focus on those conducting broader analysis on a full-time basis.

On average, bureaus reported an average of 9.6 full-time staff, three staff as half-time (50%-75%) or partial (25-50%), and six staff who could be defined as spending a minimal number of hours (less than 25%) on evidence-building activities. Staffing distributions, regardless of level of effort, were positively (right-) skewed (i.e., the median was less than the mean) due to operating units whose primary functions include providing data-intensive services across the Department. Thus, the median provides a better indication of the typical number of staff focused on evidence-building



activities in bureaus. Standard deviations—i.e., how dispersed data are compared to the mean—showed high variations in the number of individuals employed by bureaus and independent offices.

*Table 1, Staff dedicated to evidence-building (questions 9 and 10 in the Capacity Assessment survey)*

<b>Staff Level of Effort</b>	<b>Total</b>	<b>Mean</b>	<b>Median</b>	<b>Standard Deviation</b>
Full-time evidence-building role (at least 75% of time)	421	9.6	2.0	22.44
Half-time (approximately 50-75% of time)	130	3.0	0.0	7.75
Partial (approximately 25-50% of time)	134	3.0	1.0	6.04
Minimal (less than 25% of time)	244	5.5	0.5	14.50

The highest variation, however, can be seen within the numbers reported for full-time evidence-building staff. Several bureaus have functions that are inherently tied to evidence-building, such as bureaus that offer research services to other units; collecting, analyzing, and reporting on the Department’s workforce; or implementing certain legislation like the Digital Accountability and Transparency (DATA) Act. If these nine data-intensive offices are removed from the calculations for full-time staff, the total number decreases to 138 evidence-building staff, with a mean of 3.5 staff per bureau or independent office (median: 1.0; standard deviation: 4.59). This subset displays a fairly large standard deviation as 11 bureaus or independent offices reported that they do not have full-time evidence-building staff and 9 reported just one full-time staff person—continuing to illustrate the variability behind evidence-building coverage across the Department.

### Activities and Systems

To further understand how the Department’s bureaus or independent offices institutionalized evidence-building activities, the team requested that respondents answer several questions related to the type of data that are collected and analyzed.

In line with new Federal evidence-building requirements at the agency level, the Department was interested in understanding whether its bureau-level units have begun to develop their own learning agendas as a way to meet their evidence needs. As expected, this activity is at a nascent stage but shows potential for rapid growth. At the time of reporting, five respondents (11%) noted that their units have a bureau-level learning agenda that focuses on questions and evidence priorities that would specifically improve their missions. A larger proportion (25%) noted that they are in the process of developing a bureau-specific learning agenda or have something similar such as an evaluation plan. This question allowed the Department to identify opportunities to encourage bureaus and independent offices to incorporate this activity within their own infrastructures, continuing an existing process of providing evaluation plan and learning agenda background materials, technical assistance, and review of new evaluation plans and learning agendas.



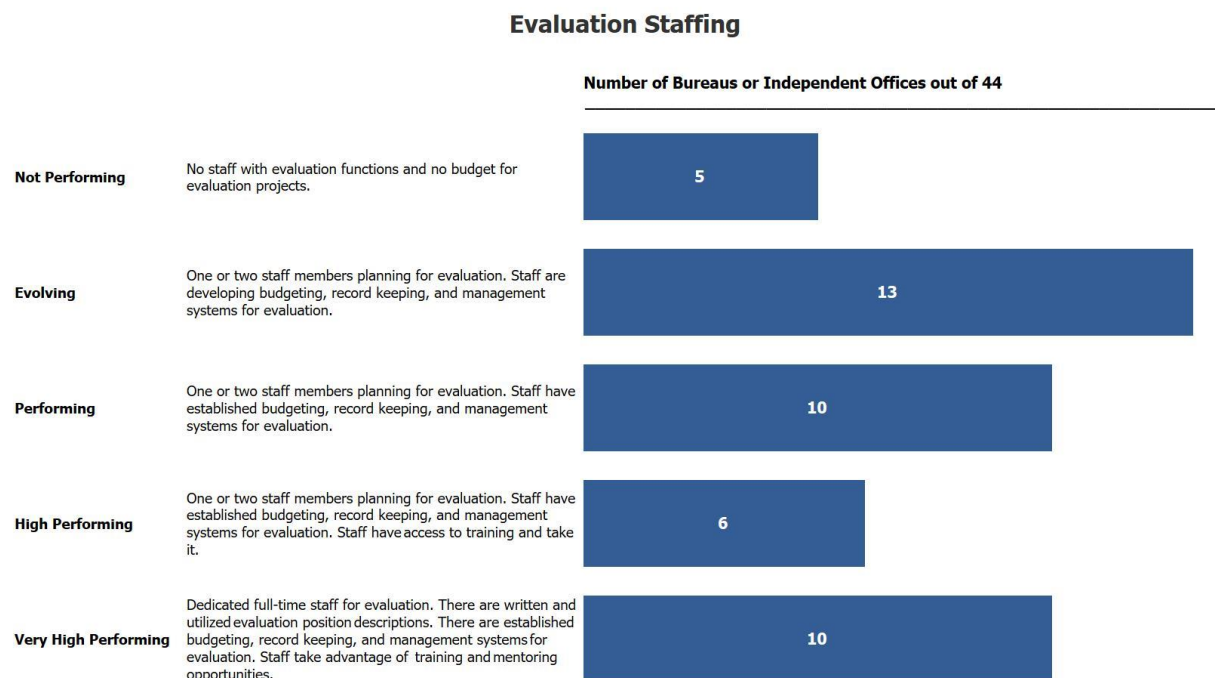
## Evaluation Coverage

Across the Department, 72% of bureaus and independent offices have commissioned external evaluations in the past three years, as reported through survey data and validated through our evaluation databases. From an analysis of survey data on evaluation activities and our evaluation database, 28% of the Department has not reported such evaluations in recent years. This figure does not necessarily identify a need. It was observed that several bureaus that have not completed evaluations have been active in conducting other research and analysis activities. Thus, the current need for evaluation would need to be further examined to understand whether it is being met.

Respondents also rated their bureau or independent office considering the categories within the maturity model. As illustrated in Figure 2, the largest proportion of respondents (13/44, 30%) assessed their bureau or independent office in the ‘evolving’ stages of evaluation staffing, meaning they have one or two staff members who are in the process of developing budgeting, record keeping, and management systems for evaluation.

In addition, the Department has a contract mechanism in place and staff dedicated to supporting bureaus through the process of designing and commissioning evaluations. Instituted in 2017, the contract vehicle features numerous slates of service providers with research expertise available to cover the diverse needs of the Department. As of August 2021, the Department has obligated \$59 million under the five-year mechanism, with \$17 million more planned.

Figure 2. Evaluation staffing maturity



Focusing on Bureau Evaluation Plans may be a way forward for increasing evaluation practice, if needed, and encouraging evidence-building in other areas like foundational fact finding. Bureau Evaluation Plans, required under Department policy, are more easily developed than learning agendas, and include fields for ongoing and future projects. As the Department increases its demand and supply for evidence-building, these databases could expand to include fields tied to evidence-building such as foundational fact finding.

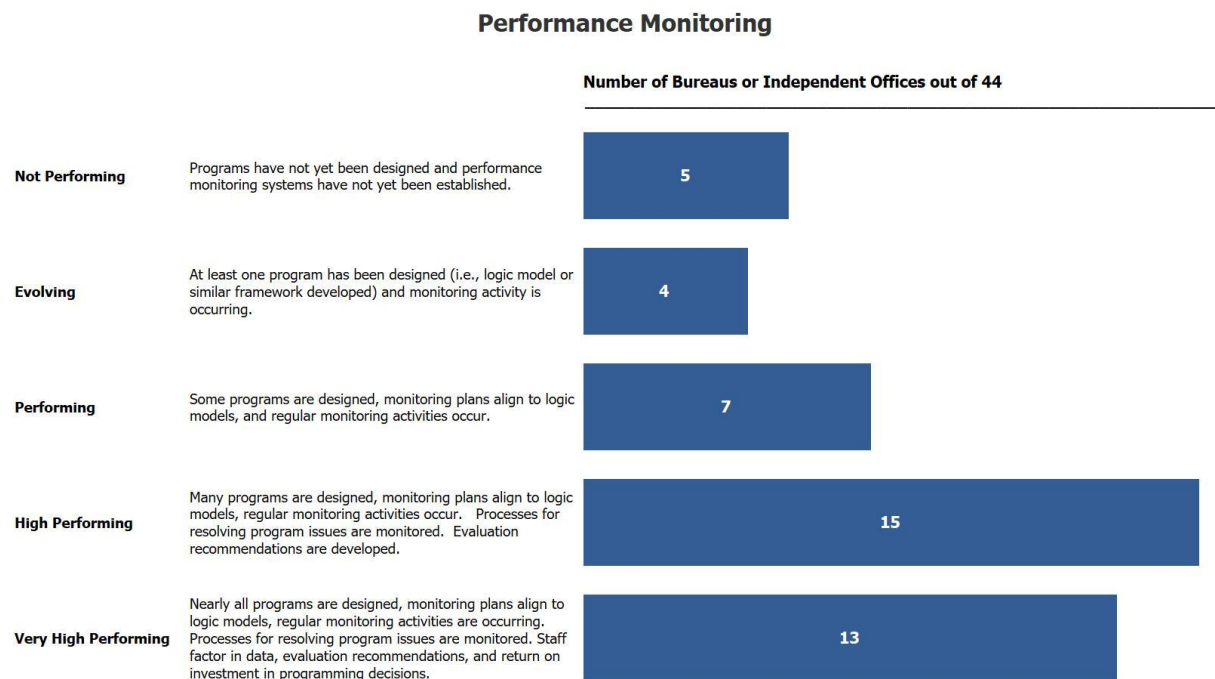
### Performance Monitoring and Program Design

As one way to understand performance monitoring and program design coverage, respondents confirmed the maturity level of these activities for their bureaus and independent offices, as noted in a maturity model, as displayed in Figure 3. Within bureaus and independent offices, evaluation continues to be closely tied to performance monitoring and program design as many evidence-building experts lead both activities. For instance, many survey respondents and focus group participants noted titles that included both duties, like Monitoring & Evaluation Specialist. When comparing the 16 bureaus and independent offices in Figure 2 that noted high or very high levels of maturity in evaluation staffing, they were more likely to confirm that their operating unit also had a high or very high level of maturity regarding performance monitoring and program design.

For the five bureaus and independent offices reporting that they have not instituted formal program design practices, work would need to be done with these units to facilitate compliance with Department policy as the 18 FAM 301.4-2 notes that operating units' program design responsibilities should include the development of situational analyses and logic models that are linked to goals, objectives, and performance monitoring activities. As the Department will be publishing new bureau strategic plans and deva performance management toolkit to accompany them, these new documents represent an opportunity for these remaining bureaus to complement their long-term strategy with program design.

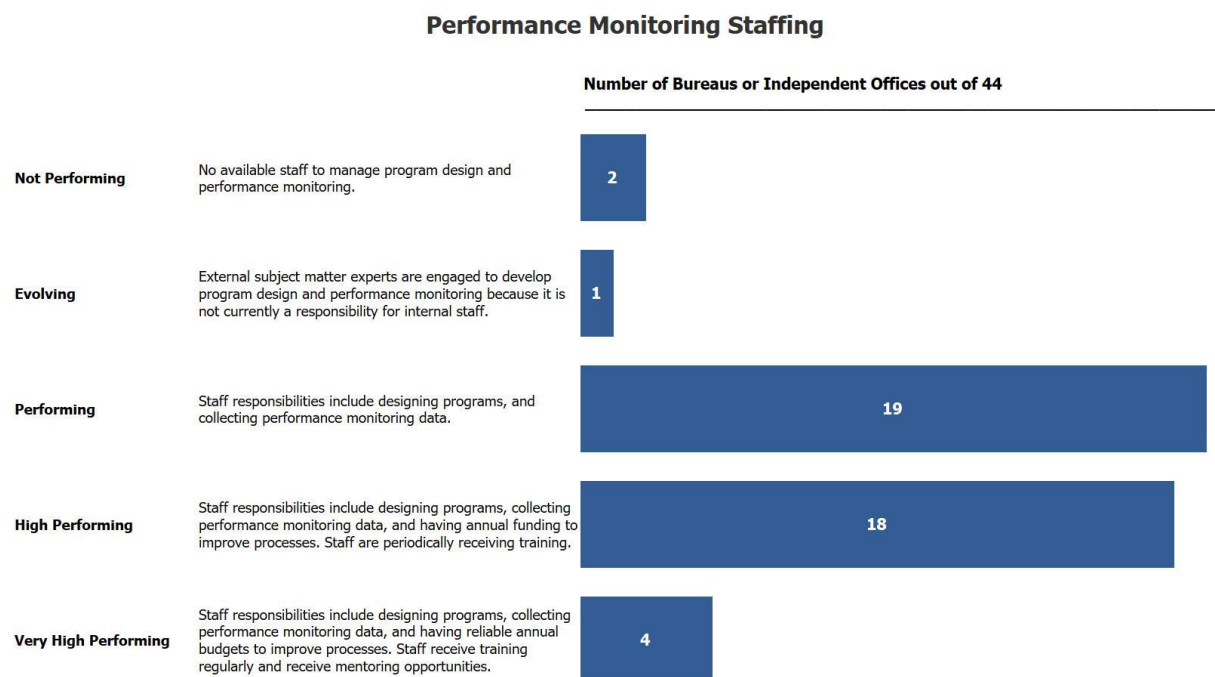


Figure 3. Performance monitoring maturity



In addition to assessing the maturity of program design and performance monitoring activities, the survey asked respondents to note their staff's completion of requirements for program design. While the majority of bureaus and independent offices (28/44, 63%) rated themselves high or very high performing in program design, we observed a decrease in the distribution for program design staffing capacity (see Figure 4). That said, 93% of respondents still noted that their bureau or independent office has incorporated program design within the day-to-day responsibilities of program staff. Staff training on program design and performance monitoring seems to be the differentiating factor across the higher levels of maturity. All 28 bureaus and independent offices that assessed their program design and performance monitoring to be high or very high also reported that staff have program design responsibilities. In many cases, as noted within surveys and focus groups, internal monitoring & evaluation specialists manage capacity building efforts in their bureaus and independent offices.

Figure 4. Performance monitoring staffing maturity



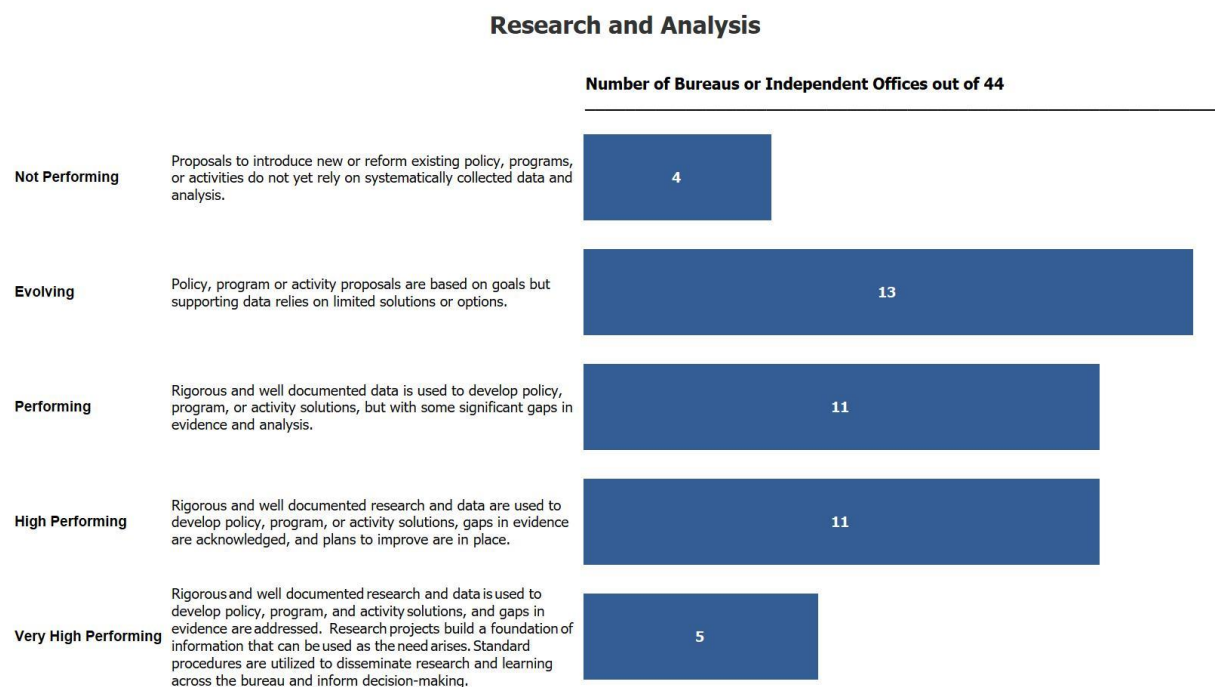
### Research and Analysis Coverage

While there tended to be a strong association between program design/performance monitoring and evaluation activities, there was a slightly weaker association between these maturity levels and research and analysis activities. As illustrated in Figure 5, 39%, or 17 out of 44 respondents placed their bureaus and independent offices in the two lowest maturity levels, which was slightly lower than the proportion noting this maturity level for evaluation staffing (18/44, 41%). Although bureaus and independent offices with higher maturity levels in evaluation, program design, and performance monitoring were still more likely to assess their research and analysis activities higher than others, there was still a mix of maturity levels spanning the lowest to highest categories in this model. This may be attributed to the varying evidence-building needs and activities across the Department.





Figure 5. Research and analysis maturity



Many bureaus and independent offices house large program units (i.e., mission-strategic activities), which require systems for program design, performance monitoring, and evaluation. Other units primarily focus on agency-operational activities, and emphasize researching and analyzing the Department’s training systems, human resources, financial compliance, or providing analytical support to other bureaus. During focus group discussions, several individuals from bureaus focusing primarily on foreign assistance noted that their staff focused on non-foreign assistance activities generally conduct less research and analysis. As the Department continues to improve evidence-building capacity across the agency, additional effort may be needed to ensure that various divisions’ evidence needs are fully addressed, regardless of activity type.

Furthermore, although the Department has systems to track program design and evaluation activities, it is difficult to fully catalogue evidence-building activities. Research and analysis activities are particularly challenging to track given they can be commissioned through a variety of mechanisms, including research grants (monitored in SAMS Domestic, a grant making database), contracts (monitored in Ariba, a purchase order database), or developed as databases or analytical systems (monitored through FedRAMP authorization to operate (ATO) certificates as well as contracts).

## Methods

In relation to the criteria of coverage (i.e., what activities are happening and who is doing them), the Department was interested in understanding the types of evidence-building activities and methods

conducted by bureaus and independent offices, as well as their views on the appropriateness of these methods in meeting their evidence needs.

In addition to assessing their levels of maturity in program design, performance monitoring, evaluation, and research and analysis, respondents noted which evidence-building activities have been consistently carried out across their bureau or independent office over the last three years (see Figure 6). In this question, respondents selected activities institutionalized and systematized within their units, not simply those considered a best practice conducted by a select few. From this question and in line with the program design and performance monitoring maturity model data, we observed that 88% of bureaus and independent offices collect performance monitoring or project indicator data.

For the most part, bureaus and independent offices conducted several performance-related activities, including managing performance data within a central repository (67%) and analyzing this performance data (74%). Bureaus also conduct other forms of program management activities required of program managers and implementing partners; these included holding periodic check-in meetings with partners and contractors (84%) or reviewing routine reporting (74%). Areas of potential growth, however, existed more within the realms of evaluation and research and analysis.

*Table 2. Most frequent evidence-building activities used by bureaus (question 12 in Capacity Assessment survey)*

<b>Bureaus Confirming Evidence-Building Activities</b>	<b>Percentage</b>
Collecting performance monitoring data/project indicator data	88%
Holding periodic check-in meetings (monthly, quarterly) with implementing partners or contractors	84%
Reviewing or assessing milestones and performance indicators related to bureau strategic plan goals and objectives	79%
Analyzing performance monitoring data/project indicator data	74%
Reviewing quarterly or other periodic reporting for funded projects	74%
Compiling project achievements/outcomes (narrative data) in a central repository or database	67%
Conducting internal research and analysis activities (conducted by internal staff)	67%
Conducting periodic data quality checks	67%
Managing performance monitoring data/project indicator data, including data input, within a central repository or database	67%
Conducting site visits/activity monitoring trips	56%
Commissioning external evaluations of your programs	53%
Conducting internal evaluation (conducted by internal staff)	47%
Commissioning external research and analysis activities	30%

Evaluation activity has continued to grow over the last several years and data from the Department’s evaluation registries suggests that most evaluations are performance or process-related (79%) evaluation work—as opposed to summative impact or ex-post. Furthermore, a cursory analysis of



registered evaluations demonstrates that some bureaus and independent offices have included questions that sought to address performance and impact ; thus, it may be necessary to review final research design documents to determine the actual evaluation type. The need to clarify evaluation type was highlighted during focus group discussions, as Department evidence-building experts observed that some evaluations are planned and implemented without serious thought to questions and, at times, a complete omission of evaluation questions. Focus group participants also noted that guidance and policies are needed to ensure that implementing partners retain data that facilitate ex-post or retrospective evaluation. In many cases, foreign assistance programs have commissioned ex-post evaluations, which has recently become an interest among Congressional leaders, but have had issues facilitating data collection due to poor data management by implementers. As the Department plans for its next phase of evidence-building capacity, it can include greater emphasis on planning for monitoring data collection and definition and structure of evaluation questions.

While nearly all bureaus noted that staff manage performance monitoring, staffing and prioritization were reported as a limitation in improving research and evaluation activities. These barriers were noted regardless of the bureau's maturity level, with 58% reporting that they have an insufficient number of evidence-building staff. Within this group, 65% noted that evidence-building staff, once hired, have insufficient time to actually conduct evidence-building activities. Research and evaluations are informed by information drawn from ongoing performance monitoring, like data calls. These challenges of competing demands for short vs. long term evidence activities are further addressed below.

During focus groups, participants noted that performance and reporting processes like the Performance/Plan Report (PPR) and collecting project indicators take the bulk of one's time, with insufficient time remaining for research and evaluation. Participants did not dismiss the value of collecting performance data but are looking for more efficient ways to do this that leaves more time for rigorous analysis via research and evaluation.

After compiling data on the types of evidence-building activities conducted by bureaus and independent offices, respondents noted the appropriateness of these methods in meeting their operational and learning needs. For the most part, respondents agreed or strongly agreed that the activities their bureaus and independent offices conducted were gathered or analyzed using appropriate methods for the task at hand. The lowest proportion in terms of appropriate methods related to compiling project achievements and outcomes (i.e., narrative data) within a central repository/database – 55% of respondents agreed and 10% strongly agreed that their bureau or independent office employs appropriate methods for this activity; the remainder selected a neutral response (28%) or disagreed (7%). There was a notable difference between the perceived appropriateness of methods utilized in internal versus external evaluations. Whereas 96% of respondents noted that external evaluations utilize appropriate methods, 79% reported this was true when evaluations were conducted by internal staff. The difference between the appropriateness of internal (83%) and external research (85%) activities, however, was less pronounced.



## Quality and Effectiveness

As previously mentioned, findings illustrate that bureaus continue to produce evaluative evidence, numbering 154 evaluations over the last three years. Department-level support for this work includes planning databases, a performance management and evaluation service contract, and trainings as listed below.

*Table 3. Resources to plan and implement evaluation activities and assist agency staff and program offices to use evaluation research and analysis approaches and data in day-to-day operations*

Activities	Type
Evaluation Management System	Planning database; registry
Evaluation Registry	Planning database; registry
Performance Measurement and Evaluation Services (PMES) IDIQ	Procurement mechanism
Strategic Planning and Performance Management (SPPM) course	Training
Managing Evaluations course	Training
Data literacy courses	Training (FSI)
External courses in evaluation, statistics, data visualization, and report design	Training

Nevertheless, bureau-level adoption of evaluations continues to be a barrier. Among survey respondents, only 43% agreed that other staff—such as program managers or policy officers—have adequate time to use evidence to inform program design and adjust operations. When it comes to identifying barriers to using rigorous methods, survey respondents primarily noted a lack of time for staff to design external activities such as writing or refining evaluation contract statements of work (66%) and an insufficient number of staff focused on evidence-building activities (57%). As previously noted, when evidence-building staff are hired, they noted that much of their day-to-day activities were related to performance monitoring rather than research and evaluation.

A number of bureaus and independent offices (45%) also noted a lack of funding to conduct activities such as external evaluation. This barrier was noted during focus groups, as participants asserted bureaus and independent offices have few internal incentives or desire to conduct evaluations unless they are directed to do so externally, such as through funding allocations. These responses point to the need for greater focus on evaluations from within many bureaus' program offices and senior leadership, as related funding allocations are generally driven by internal priorities within each bureau, rather than externally. As a result, the Department refined a learning agenda question to focus on how expectations for evaluation are set and how evaluation information is used.

*Table 4. Most frequently selected barriers for using rigorous methods (question 23 in Capacity Assessment survey)*

Barriers for Rigorous Methods	Percentage
Insufficient time for staff to conduct evidence-building activities	66%
Insufficient number of staff focused on evidence-building activities	57%



Insufficient time for staff to design external evidence-building activities (e.g., writing or refining evaluation contract statements of work)	45%
Lack of funding for evidence-building activities (e.g., external evaluations)	45%
Insufficient time for staff to manage external evidence-building activities (e.g., external evaluations, external research)	39%
Lack of appropriate skills to conduct evidence-building activities	39%
Lack of demand from stakeholders	32%
Lack of appropriate knowledge to manage evidence-building activities	20%
Other (please specify)	16%

One related theme to emerge during focus group discussions was the uneven distribution of evidence-building activities within a bureau or independent office. Within survey comments and focus groups many respondents mentioned that obtaining data at the sub-organizational level of bureau or independent office may hide pockets of maturity. Respondents mentioned that evidence-building activities are generally uneven because many bureaus and independent offices focus on their own specialties. More specifically, while particular divisions or units within a bureau or independent office, especially those responsible for managing program funds, may have robust evidence-building systems, others, like policy offices or operational units, may fall more within a nascent stage.

In addition to assessing whether the methods utilized were appropriate, respondents were asked to rate the quality of their evidence-building activities. This question sought to understand internal experts' views on quality and to associate it with other quality markers that have been collected over the last several years. Across these assessments, the distribution of scores were lower for performance-related activities like collecting performance and project indicator data, when compared to activities such as research and evaluation. For performance-related activities, most respondents noted fair (i.e., fragmented planning and operationalizing) or good (i.e., sufficient) quality evidence. Similar to data on methods, respondents noted similarities in the quality of external (85%) and internal (85%) research activities (i.e., those ranking quality as good, very good or excellent); however, views on the quality of external evaluations (87%) exceeded those on internal evaluation (74%).

This self-assessment of quality was supplemented with data from a previous independent meta-evaluation of the Department's evaluation activities.<sup>1</sup> In line with the findings from our capacity assessment survey, the previous evaluation found that respondents reported that evaluation deliverables were generally fair or good. In this evaluation, an assessment was then conducted on the content of evaluations based on a number of standard criteria—including description of methodology, findings draw on data collection methods, report answers all evaluation questions.

<sup>1</sup> See Department of State *Examination of Foreign Assistance Evaluation Efforts at the Department of State: A Combined Performance and Meta-Evaluation* (2019), available at <https://www.state.gov/wp-content/uploads/2018/12/2018-Examination-of-Foreign-Assistance-Evaluation-Efforts-at-the-Department-of-State-A-Combined-Performance-and-Meta-Evaluation-.pdf>. See specifically pages 9-14 for findings related to the quality of evaluation reports.



Several areas—including report structure, methodology, conclusions, and recommendations—were assessed as good or as meeting 75% or more of the quality indicators, for the majority of evaluations reviewed.

However, there were several areas for improvement. The following areas across the dataset were found to be of fair quality (i.e., meeting 50-75% of the quality indicators): clearly stated objectives and audience for the evaluation (which could also be seen as an issue with evaluation statements of work), developing quality evaluation questions (potentially also an issue with design), and presenting findings in a way that uniformly draws on data collection methods as well as discussing possible alternatives.

A recent informal review of evaluations by Department staff conducted after 2018 using the same criteria used in the meta-evaluation shows similar trends. While most of the Department's recent evaluations include findings related to the evaluation questions, it is uncommon to see alternative explanations examined at length within evaluation reports. It is possible that this stems from the high proportion of evaluations focusing on process and performance. Most evaluation reports did not incorporate more rigorous designs that could be used to compare results like process tracing, and comparative case studies. As noted in the 2018 meta-evaluation, quality issues often translate to issues of utility, such as unactionable findings and recommendations.

To understand how the use of more rigorous methods may relate to evidence quality, we correlated these data to understand if there were relationships between these two areas. For all but two types of evidence—specifically, managing performance / project data within a database, and compiling project achievements—there was a moderate (correlation coefficient,  $r = 0.4-0.59$ ) to strong ( $r = 0.6-0.79$ ) positive relationship. Additionally, we correlated internal experts' assessments on the quality and utility of evidence as we were interested in understanding whether evidence quality may counter barriers such as insufficient staff time. With regard to quality and utility, moderate to strong positive correlations were observed for six of the 13 evidence types noted by respondents. Correlations were fairly weak for activities conducted by program staff, such as conducting periodic data quality checks ( $r = 0.14$ ), conducting site visits/activity monitoring trips ( $r = -0.03$ ), reviewing quarterly/periodic reports for funded projects ( $r = 0.35$ ), and compiling project achievements ( $r = 0.26$ ). It should also be noted that higher assessments of quality did not relate to higher assessments of utility for internal research ( $r = 0.10$ ) and internal evaluation ( $r = 0.04$ ). This was not the case, however, with external research ( $r = 0.55$ ) and evaluation ( $r = 0.65$ ) as we observed moderate to strong correlations between quality and utility. This finding was somewhat unexpected as the assessments on the appropriateness of methods and quality of internal and external research activities were nearly identical.

While unexpected as a finding, it is not entirely surprising as respondents noted that internal staff generally have insufficient time to employ more rigorous methods in their evidence-building activities. Regarding quality and utility, it seems that there is significant work to do in improving the type of evidence like performance indicators collected by program staff, as well as exploring how the utility of internal evaluation can be improved so they are of use to Department decision-makers. Issues concerning evidence quality may be further substantiated in the statistics provided by





respondents on the available skills within their bureaus and independent offices. When asked whether staff in their operating units had particular research and evaluation skills, fundamental techniques were present within a majority of units like fundamental statistics (65% agreed) and fundamental qualitative analysis such as content or thematic analysis (57%). More nuanced techniques—such as advanced statistics including inferential and Bayesian techniques (28%), experimental or quasi-experimental design (33%), and advanced qualitative analysis such as narrative inquiry or process tracing (43%)—were less likely to be observed.

### Disseminating Good Practices and Findings

The survey asked respondents whether they disseminated their findings and with which audiences these were most often shared. Over 80% of respondents reported that their bureaus and independent offices shared findings internally with program officers (84%), leadership such as deputy directors and directors (88%), and front office staff such as Deputy Assistant Secretaries and Assistant Secretaries (86%). The majority (65%) also noted sharing evidence with other bureaus or independent offices within the Department. Sharing outside the Department, however, was less likely as only 42% reported disseminating evidence with other federal agencies and 35% with Congress. Although implementing partners (i.e., grantees and contractors) often provide evidence to the Department in the form of project data and special reports, bureaus and independent offices do not as frequently return the favor, with only 33% sharing evidence with this stakeholder group. Evidence was shared less frequently with academic researchers (12%) and the public (19%).

With regard to how the Department disseminates good practices, the team sought better understanding of respondents' perspectives. Respondents were asked which Department resources they considered the most useful for their evidence-building activities. Approximately 36% responded that policies and processes covering strategic planning and performance monitoring and evaluation were most useful, ranking this option as their first choice. During focus group discussions and within open-ended survey questions, respondents also mentioned the need for offering standard tools and activities (topics of interest are included in the following section) that can be used by all bureaus and independent offices. This finding was similar to one from the 2018 performance and meta evaluation of the Department's foreign assistance evaluation efforts. Since then, the Department designed and produced coursework on strategic planning and performance management and published a toolkit on performance management, which includes preparing for evaluation. Given this progress, additional emphasis could be placed on reminding bureau evidence-building staff that Department-specific resources are available on an internal website.

Approximately 25% of respondents noted that online resources such as the Department's Managing for Results website, Performance Monitoring and Evaluation Services IDIQ contract vehicle, Program Design and Performance Management toolkit, and Department-sponsored training including the Managing Evaluations course and Strategic Planning and Performance Management, were most useful. The Department's communities of practice for those interested in evaluation and program design were only ranked first by 7% of respondents, with the majority of respondents (66%) selecting it as their third or fourth choice or noting that it was not applicable to their bureau or independent office. This ranking may be entirely consistent with the immediate value that training



and contracting tools provide when designing and commissioning an evaluation, as compared to communities of practice, which tend to have a longer-term learning and networking value.

### Day-to-Day Operations and Learning Needs

In addition to the dissemination of good practices, the Department sought to understand how it can further assist agency staff and program offices to use evaluation research and analysis approaches. As such, respondents were provided a list and asked to select the types of information and tools that would support their evidence-building and day-to-day operations.

As illustrated in Figure 8, evidence-building experts were most interested in accessing tools that would support activities in data visualization, data analysis, and data management. Although an extensive list of software packages have been cleared for use within the Department, focus group participants noted difficulties in procuring these tools. Furthermore, they noted that while these tools may be available, it does not ensure use as the time to conduct internal evidence-building activities may be constrained by other activities such as extensive performance monitoring and reporting requirements. Near the top of this list of tools in demand by participants (see Figure 8) is funding, with 57% of respondents noting that additional funding is needed for evidence-building activities. One theme which arose during focus groups was the need for the Department to place additional emphasis in coordinating evidence-building and learning activities among bureaus and independent offices so that they may learn from one another and share data.

Evidence-building staff identified pockets of expertise and noted that they were interested in understanding how other bureaus and independent offices conduct evidence-building work. Some expressed interest in jointly-funded evaluations on similar topical areas. In this regard, opportunities remain in coordinating and facilitating collaboration among bureaus. With regard to training, respondents expressed interest in learning from their counterparts, with 50% expressing interest in being trained by a Department advisor who could provide good practices on evidence-building in a way that is contextualized to the Department's operations. Another option is to more consistently include evaluation requirements into the program awards (whether grants or contracts), so that they are planned for at the beginning of a program cycle.



Table 5. Most frequently requested evidence-building information or tools (question 30 in Capacity Assessment survey)

<b>Types of Evidence-Building Information or Tools Requested by Respondents</b>	<b>Percentage</b>
Tools for data visualization	61%
Additional funding for research, M&E, and learning	59%
Tools for data analysis	59%
Access to data from within the Department	57%
Access to an advisor who can assess methodological or analytical questions	55%
Short guidance notes (e.g., step-by-step instructions, how to notes) on evidence generation, management, dissemination, and use	55%
Tools for data management	55%
Access to data from external sources	50%
Direct training/technical assistance from a Department evaluation advisor	50%
Tools for sharing and disseminating evidence	48%
Direct training/technical assistance from external experts (e.g., academics, researchers)	39%
Additional guidance on commissioning and procuring evidence-building activities	32%
Additional policy guidance (e.g., 18FAM300)	27%
Other (please specify)	27%
I don't know	2%

At the moment, several bureaus have developed data systems for their own use, and the Department continues to build standardized systems for enterprise-wide data sharing. As many bureaus have developed their own systems, respondents saw enterprise solutions as a way to leverage existing data, standardizing processes based on what has worked, and further facilitating collaboration. This need for data sharing was selected among 54% of survey respondents and noted it as just one way to develop and sustain a culture of evidence across the Department. Additionally, a majority of respondents (52%) also noted the need to standardize evidence-building activities through guidance documents, such as step-by-step instructions and how-to notes on good practices for evidence generation.

### Capacity-Building Activities

As training and mentorship opportunities are vital components for capacity building and continued improvement, the survey asked respondents to assess their bureau or independent offices' support for training for direct hires interested in evidence-building activities. Of the 44 respondents, 40 (91%) noted that their unit currently supports training for direct hire staff. However, within follow-up questions and comments made during focus groups, bureau evidence-building experts mentioned several obstacles to training, even if their bureau currently supported it. These obstacles often relate to a lack of time due to work requirements. Several respondents, within surveys and during focus

groups, also mentioned varying priorities as those working on Department policy may not have time or funding for evidence-building training.

In terms of program design, several focus group participants mentioned that program officers may not hold the requisite expertise to systematically utilize evidence when designing programs, in turn affecting the level of interest in research and evaluation. Additionally, a perceived lack of training for official roles within bureaus and independent offices, such as Bureau Planners or Bureau Evaluation Coordinators, appeared as a limitation in terms of the quality and effectiveness of evidence-building activities. This was an interesting perspective as the Department does offer strategic planning, program design and evaluation management training. Additional inquiry may be necessary here to understand the roots of this, which could include frequency of course offerings, the need to orient new staff to training options, the importance of emphasizing external training opportunities, and options for more specialized training designed for specific programs.

While the Department continues to hire more evidence-building staff, the findings noted that data and evidence literacy must be improved across the agency to ensure that those who plan and implement programs understand and utilize appropriate evaluation, research, and analysis tools. Although staff have limited time to utilize evidence, focus group participants added that more training for staff and leadership could strengthen a baseline level of understanding of how evidence should be utilized across various levels. To address these findings, the Department is participating in a data analyst hiring initiative and is increasing focus on program design and performance management as well as longer term research and analysis.

## Independence

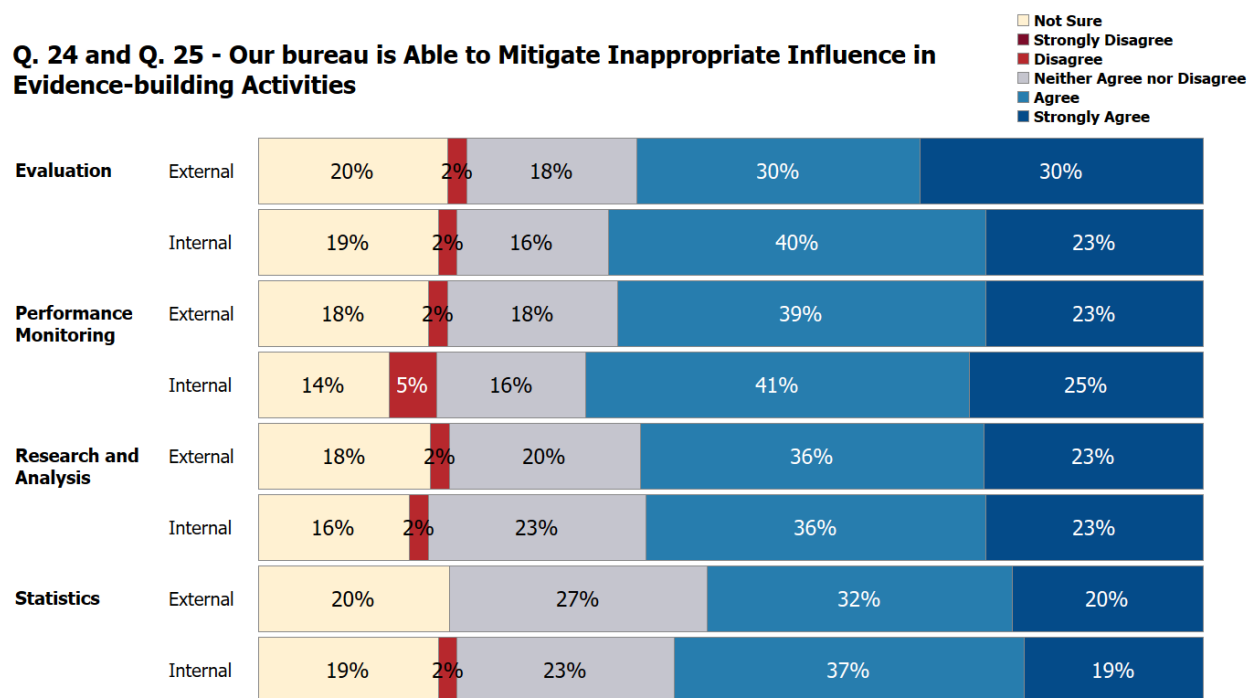
In an effort to understand how the Department's evidence-building leaders view the concept of independence, we asked about their views on whether their bureau or independent office is able to mitigate "inappropriate" influence in both internal and external activities. To inform M-20-12's focus on "bias and inappropriate influence," this assessment defined mitigating "inappropriate influence" as the ability to systematically and fairly consider evidence in an objective manner.

As illustrated in Figure 9, for all evidence-building activities, more than half of respondents expressed that they either agreed or strongly agreed with the statement, "Our bureau is able to mitigate inappropriate influence in terms of maintaining objectivity, impartiality, and professional judgment." Respondents demonstrated the most confidence in ability to mitigate inappropriate influence via internal performance monitoring activities (66% agreed or strongly agreed.) External evidence-building activities related to statistics, for instance data collection and processing conducted by contractors or grantees, held the lowest proportion of respondents who agreed (32%) or strongly agreed (20%) with this statement: The proportion of respondents noting their agreement in mitigating inappropriate influence over internally-generated statistics was similar (56%). These findings may relate to the respondents working definition of "statistics" as method of analysis, rather than as distinct 'statistical programs.'



Consistent with the definitions in OMB memo M 20-12, when bureaus and independent offices commission evidence-building activities from external sources, the evaluators are vetted on their skills and they must also sign agreements acknowledging and mitigating conflicts of interest.

Figure 6. Bureau ability to mitigate inappropriate influence (questions 24 and 25 in Capacity Assessment survey)



## CONCLUSIONS

### Agency Capacity

The Department of State continues to make progress since the implementation of the first policies and procedures to advance evidence. However, although the Department provides policy and guidance on evidence-building functions, it is unclear whether this information flows effectively to those who may not be responsible for evidence-building activities, thereby creating a silo of knowledge and maturity amongst evidence-building staff. Furthermore, with the increasing focus and statutory mandates surrounding evidence collection, building, and use, Department staff with these responsibilities face competing demand on their time and resources, generally emphasizing performance and reporting activities over research and evaluation. In addition, evidence that is generated is not always used or shared widely so other bureaus can learn from it.

Standardized procedures, along with tailored evidence building tools, enterprise-wide data sharing, and systems by which existing expertise can be leveraged readily would help bureaus expand their knowledge base for evidence building. Developing and sustaining a culture of evidence building across the Department will require transparency and enterprise-wide data sharing capabilities as well as consistent and repeated communication about existing systems, tools, guidance, and policy.



## ISSUES FOR CONSIDERATION

**Bolster the evidence-building culture in the Department through targeted hiring, including of specialists in data analysis, evaluation, research, and learning and by ensuring that performance is measured against capacity to integrate data and learning into strategic planning:** Evidence-building staff responsibilities often focus on conducting performance monitoring, rather than the longer-term research and evaluation activities. As the Department develops guidance and standards on evidence-building, there is an opportunity to develop hiring and workforce development standards. This will help ensure that the Department's professionals have the necessary skills to balance performance monitoring and evaluation needs.

**Strengthen evaluation presence in Department-wide processes:** Department-wide planning processes including resource strategic reviews, Joint Strategic Plan (JSP), Annual Performance Report (APR), and the Performance Plan and Report (PPR) could include sections that draw upon bureau learning agendas and evaluation plans. Processes should encourage bureaus to explore research and evaluation findings and should include an expectation that bureaus and independent office will summarize this information each year when discussing large-scale programs.

**Promote evidence literacy across the Department and connect bureaus with the Center for Analytics and other data analytic tools in the Department:** From survey and focus group findings, the evaluation found that evidence-building experts thought that Department leaders and working-level staff need more time and better skills to integrate evidence to inform policies and programs. While the Department currently administers several courses, such as the week-long Strategic Planning and Performance Management training, these activities could be more heavily marketed to non-evidence-building staff. There are a range of analytic tools and capacities in bureaus such as the Bureau for Conflict and Stabilization Operations (CSO), the Bureau for Intelligence and Research (INR) and the Office of Policy Planning (S/P).

**Integrate other forms of evidence within registries and invest in analyses of evaluations with the goal of building information and data that a wider audience across the Department can access:** Currently, the Department tracks evaluations through the use of two databases. However, there is no formal way to account for the breadth of evidence as other evidence-building activities like foundational fact finding can be funded through various mechanisms like grants, contracts, and interagency agreements, and are tracked in separate systems. As the Department continues to build systems to standardize how narrative data can be stored, additional work can be done to ensure that project achievements are collected in systematic ways, rather than relying on ad hoc anecdotes that may arise during periodic check-in calls. The Department's internal data catalog/inventory, managed by Center for Analytics, may also be a useful resource. In addition, State Department offices using data should determine how to integrate other data from outside of government on relevant issues such as climate, public health, and conflict.

**Reinvigorate professional development and training opportunities for research, evaluation, and learning staff within the Department to ensure familiarity with data and analytics skills:** The Evaluation Community of Practice provided regular education from external presenters,





knowledge exchange from bureau presentations, and access to the Department's internal expertise. Along with trainings, workshops, and established tools, the Community of Practice provided a forum for learning, exchange, updates, and discussion among community members. Reinvigorating it, looking at how best it can serve the community, and considering other ways to contribute to dialogue, collaboration, and learning could help strengthen evidence-building practice at the Department.

**Share evidence building activities across bureaus, including data, analysis, and evaluations collected in some bureaus with relevance or adaptative potential to other bureaus:**

Participants proposed getting access to systems that allow bureaus to see the top-line areas that others are supporting. This could be accomplished through existing systems (like SAMS Domestic-a grant making database) by adding reference points to the Standardized Program Structure and Definitions (SPSD) numbers and categories, which could then provide a systematic way to codify projects. Considering any procurement concerns, data would need to be accessible to all relevant users. For instance, before a bureau implements a project on rule of law, a bureau or independent office would check with other relevant bureaus to see what has been funded or who is responsible for managing those types of projects.

**Recognize evidence-building funding across the Department through budget formulation and operational plans as well as through senior leader prioritization:** Currently, while some bureaus set aside a percentage of funds for evidence-building activities, others rely on unspent funds near the end of the fiscal year. Bureaus and programs should consider more transparently identifying and tracking evidence-building activities in their annual budgets.



---

# APPENDICES

---



## APPENDIX A: METHODOLOGY

*Table 1. Capacity Assessment working group webinars*

Date of working group session	Purpose
Session 1: December 2020	<ul style="list-style-type: none"> <li>• Review purpose of the Capacity Assessment and how the Department intends to develop and use it</li> <li>• Review the Maturity Model, and provide suggestions on enhancing it</li> </ul>
Session 2: January 2021	<ul style="list-style-type: none"> <li>• Review revisions to domain definitions and maturity model</li> <li>• Describe initial plans for data collection tools</li> <li>• Discuss strategy for survey launch</li> </ul>
Session 3: Planned, October 2021	<ul style="list-style-type: none"> <li>• Review data and recommendations</li> <li>• Solicit stakeholder feedback</li> </ul>

### Maturity Model

The maturity model demonstrates potential growth trajectories for bureaus and independent offices within the domains of performance monitoring, evaluation, statistics, and research and analysis. The model adopts the domains identified in the Evidence Act, evaluation, statistics, research and analysis and adds a domain for performance monitoring, which has been a focus area for the Department. It also includes a sub-domain that focuses on staffing and training in order to understand the degree to which bureaus and independent offices are investing in the staffing to do this work. Each domain contains five levels of maturity, from “Not Performing” to “Very High Performing.” Because the Department did not have established definitions with which to explore statistics, research and analysis, definitions were developed in collaboration with stakeholders to provide respondents a way to focus their thinking and answers.

The maturity model is the product of multiple layers of expertise from throughout the Department. During working group sessions, Department subject matter experts refined a draft model, providing feedback on type and number of domains and sub-domains as well as appropriateness of capacity levels. Leadership then provided feedback on the next version of the model, before it was integrated into the final survey tool.

### Desk Review of Existing Data

#### Process

The team identified existing data and mapped it to maturity model domains and criteria in order to categorize it for future data analysis. The datasets include information on staff training, program design and policy implementation data, evaluation services sourced through contract mechanisms, evaluation data and quality assessments of performance indicator usage, and data skill maturity



information supplied by the Chief Data Officer. These existing data were analyzed and aggregated to contribute to the understanding of the Department’s baseline capacity.

## Survey Design

### Conceptual Map and Measurement Priorities

The team consulted legislation, OMB guidance, and other resources to develop a conceptual map that related requirements to data, supported the creation of research tools, and focused the capacity assessment analysis. The team then explored whether existing data could answer questions posed by the Evidence Act, and prioritized data gaps for the survey.

The end result of this design work was a maturity model that allowed respondents to identify their current capacity, and a complete survey tool that explored the dynamics behind current capacity and prospects for growth. Focus groups also collected perspectives on current capacity of the foreign assistance community in evidence-building coverage, effectiveness, quality, independence, and methods.

### Process

The conceptual model allowed the team to develop the survey and focus groups questions. This tool helped the team ensure that it accounted for and researched Evidence Act criteria.

*Table 2. Evidence Act criteria linkage to Capacity Assessment survey questions*

<b>Criteria: Evidence Act elements / framing</b>	<b>Survey items and existing data (associated domain)</b>
<b>Coverage:</b> What is happening, where is it happening and who is doing it?	Questions on evidence-building offices and staffing levels, maintenance of a learning agenda, skillsets and training (all domains)
	Existing data on Department-wide procurement mechanisms and activities in support of evidence-building (all domains)
	Questions on evaluations produced, evaluation staffing, planning, and training (evaluation)
	Existing data from evaluation registries (evaluation)
	Questions on program design and performance monitoring practice, staffing, and training (performance monitoring)
	Question on extent to which research and data is integrated into policy (research and analysis)
<b>Effectiveness:</b> Are the activities meeting their intended outcomes, including serving the needs of stakeholders and being disseminated?	Questions on dissemination of data, usefulness of data, capacity building utility, additional tool preference (all domains)



Criteria: Evidence Act elements / framing	Survey items and existing data (associated domain)
	Existing data from 2018 meta-evaluation of foreign assistance evaluations (evaluation) Question on extent to which research and data is integrated into policy (research and analysis)
<p><b>Quality:</b> Do activities use appropriate methods and with the necessary level of rigor; and are the data used of high quality with respect to utility, objectivity, and integrity?</p>	<p>Questions on sufficient time to use evidence, barriers to using rigorous methods, utility of capacity building and preferred tools (all domains)</p> <p>Existing data from 2018 meta-evaluation of foreign assistance evaluations (evaluation)</p> <p>Question on evaluation staffing, planning, and training (evaluation)</p> <p>Questions on program design and performance monitoring practice, staffing, and training (performance monitoring)</p> <p>Question on extent to which research and data is integrated into policy (research and analysis)</p>
<p><b>Methods:</b> What are the methods being used for these activities?</p>	<p>Questions on appropriate methods applied, barriers to using rigorous methods, skills present in bureaus (all domains)</p> <p>Existing data from evaluation registries (evaluation)</p>
<p><b>Independence:</b> To what extent are the activities being carried out free from bias and inappropriate influence?</p>	<p>Questions on mitigating inappropriate influence in internal and external evidence-building activities. (all domains)</p> <p>Existing 2018 meta-evaluation of foreign assistance evaluation (evaluation)</p>

The Department’s Evidence Act team, made up of staff with experience in research, evaluation, and performance monitoring, developed questions related to maturity model domains and Evidence Act criteria (coverage, quality, effectiveness, methods, and independence).

**Procedures**

**Data Collection**

Based on the Department’s operating structure, the Evidence Act team selected bureaus and independent offices as the unit of analysis for the capacity assessment with each bureau or independent office responsible for completing just one survey. The team identified both evidence-building staff and leadership in advance of survey launch and sent the survey simultaneously to increase survey completion. Leadership within the Office of Foreign Assistance and Bureau of



Budget and Planning also sent emails to Front Office leadership to ensure respondents were aware of the survey and deadline for completion.

Surveys went to designated recipients at 44 bureaus and independent offices selected within the sample. A number of units, such as advisory commissions and special representatives' offices, did not fall within that group of 44 as they did not meet the criteria of our target population due to their small size and scope, and the fact that they use data from other units that provide analytical and research services.

Of the 44 units that received the survey, 100% responded. The team credits this response rate to the communication structure and extensions provided for completion. Respondents were well-acquainted with their bureau or independent office, the Department, and represented senior level leadership. Evidence-building experts selected to respond on behalf of their units included individuals with titles ranging from Monitoring and Evaluation Specialists to Management Analysts to Directors and Division Chiefs. Of the 44 respondents, 89% worked in the Department for 4 or more years, with 59% of respondents serving in the Department for more than 11 years at the time the survey was taken. Additionally, 59% of the respondents identified as Bureau Evaluation Coordinators, which is a formalized responsibility specified in the Department's policy on program and project design, monitoring, and evaluation—as specified in the Foreign Assistance Manual (FAM). Furthermore, 52% identified their role as a Bureau Planner, which holds responsibilities for a unit's strategic planning, resource management, and performance reporting.

### Data Analysis

After exporting the survey data, the team computed descriptive statistics, as well as a series of cross-tabulations and correlation coefficients. Using the Department's maturity model, we built a snapshot of Department capacity illustrating its level of strength in performance monitoring, performance monitoring staffing and resources, evaluation, evaluation staffing and resources, statistics, and research and analysis

### Limitations and Mitigation

Although the team determined that organizational units provided the most appropriate unit of analysis, there were limitations within this sampling structure that we attempted to mitigate. Sending one unique survey link to an appropriate point of contact could bias a unit's response as answers could potentially only represent the thoughts of the one individual. To mitigate this limitation, we provided the survey to working-level staff and leadership in two formats—Microsoft Word and Adobe PDF—so they could more easily distribute surveys across their bureaus or independent office to collate responses. Additionally, points of contact were required to obtain clearance of their official survey submission at their appropriate leadership level. Extending the window of time for survey completion allowed many points of contact to check in with the various sub-organizational units across their bureau or independent office.





## APPENDIX B: FY 2023 ANNUAL EVALUATION PLAN

Table 1. List of Evaluations and their Alignment to the Department’s Learning Agenda Priority Questions

Learning Agenda Question	Evaluation	Timing
<b>Question 1: How can the State Department improve the effectiveness of its diplomatic interventions to better advance foreign policy objectives?</b>	** Evaluation of Brazil’s Youth Ambassadors and English Immersion Camp	Completed in FY 2021. Results and findings of the evaluation will inform the public diplomacy team in Brasilia with their strategic planning and program design efforts and assist in the follow up research projects.
	* Expo 2020 Dubai Evaluation	October 2021 – June 2022
	YSEALI Regional Workshop Participation Evaluation	November 2021 – January 2023
<b>Question 2: How can the Department improve the effectiveness and sustainability of its foreign assistance efforts?</b>	Global Drug Demand Reduction Impact Evaluation	January 2021 – December 2025
	* International Narcotics and Law Enforcement Affairs Program Evaluation in the Central African Republic	June 2021 – December 2021
	* Performance Evaluation of the U.S.-Jamaica Child Protection Compact Partnership Evaluation	September 2018 – June 2022
	* U.S-Peru Child Protection Compact Partnership Evaluation	September 2017 – June 2022
	Practical Evaluations and Exercises (PE2)	October 2020 – September 2024
	Evaluation of Counterterrorism Programs	September 2020 – September 2023
	Prisons-Related Program Design, Monitoring and Evaluation Support to the Counterterrorism Bureau	October 2020 – September 2023
	* Fundamental Freedoms Funds Research, Evaluation and Learning Initiative	October 2018 – April 2022



Learning Agenda Question	Evaluation	Timing
	* Midterm Evaluation of the South Asia Small Grants Program	October 2021 – March 2022
	Multi-year Evaluation of the Tomorrow’s Leaders (TL) Scholarship Program	March 2022 – April 2024
	* * Effective Community Organizing and Mobilization	Ongoing, September 2020 – September 2022
	* * Outcome Evaluation on Implementation of Child Drug Use Protocols in India	Ongoing, August 2018 – September 2023
	* Bosnia DemCom Evaluation	November 2021 – May 2022
<b>Question 3: How can the Department’s tools best address the climate crisis?</b>	* Private Investment for Enhanced Resilience (PIER) Evaluation	January – May 2022
<b>Question 4: How can the Department better respond to unpredictable international events and emergencies such as global pandemics?</b>	* Evaluation of Population, Refugees, and Migration-Funded Mental Health and Psychosocial Support Services for Refugees	September 2021 – July 2022
	Evaluation of Livelihoods Support to Syrian Refugees	September 2022 – July 2023
	Evaluation of Population, Refugees and Migration-Supported Initiatives in Accountability to Affected Populations	September 2022 – July 2023
	* * Evaluation of Protection of Refugee Youth in Urban Areas in Africa	Ongoing, September 2021 – July 2022
<b>Question 5: How should the Department confront the rise of global disinformation and its negative effects on the security and prosperity of the United States?</b>	* EUR/ACE Media Literacy Program (Eastern Europe and Eurasia) Evaluation	September 2020 – February 2022
	* EUR/PPD Media Literacy Training Evaluation	March 2021 – December 2022 (TBD pending COVID-19)
<b>Question 6: How can the Department balance customer service expectations with national security and cost-effectiveness to provide a better customer service experience to</b>	* Global Support Service (GSS) Evaluation	March 2024 – July 2024



Learning Agenda Question	Evaluation	Timing
U.S. citizens, and to foreign nationals seeking visas?		
<b>Question 7: How can the Department more effectively analyze and manage risks to promote a safe and secure working environment for staff and partners?</b>	Office of Information Security (DS/SI/IS) Process Evaluation	January 2021 – October 2022
<b>Question 8: How can the Department utilize performance management and evaluation data and data systems to improve decision-making?</b>		

\* Evaluations not part of FY 2023 AEP that are significant and support Agency learning.

\*\* Evaluations from FY 2022 AEP and their status



## APPENDIX C: STAKEHOLDER ENGAGEMENT

<b>Stakeholder Engagement</b>	<b>Date</b>
Evidence Act Orientation	July 7, 2020
CA Working Group Session 1	December 9, 2020
CA Working Group Session 2	January 27, 2021
Cognitive Interviews	February 25, 2021
Capacity Assessment Survey Opened	April 14, 2021
Capacity Assessment Survey Closed	May 18, 2021
CA Focus Group 1	July 21, 2021
CA Focus Group 2	July 25, 2021
CA Working Group Session 3	Fall 2021



## APPENDIX D: MATURITY MODEL

Domain	Not Performing	Evolving	Performing	High Performing	Very High Performing
Performance monitoring	Program(s) have not yet been designed and performance monitoring systems have not yet been established.	At least one program has been designed (i.e., logic model or similar framework developed) and monitoring activity is occurring.	Some programs are designed, monitoring plans align to logic models and regular monitoring activities occur.	Many programs are designed, monitoring plans align to logic models, regular monitoring activities occur. Processes for resolving program issues are monitored. Evaluation recommendations are developed.	Nearly all programs are designed, monitoring plans align to logic models, regular monitoring activities are occurring. Processes for resolving program issues are monitored. Staff factor in data, evaluation recommendations, and return on investment in programming decisions.
Performance monitoring staffing, training, and funding	No available staff to manage program design and performance monitoring	External subject matter experts are engaged to develop program design and performance monitoring because it is not a responsibility for internal staff.	Staff responsibilities include designing programs and collecting performance monitoring data.	Staff responsibilities include designing programs, collecting performance monitoring data, and having annual funding to improve processes. Staff are periodically receiving training.	Staff responsibilities include designing programs, collecting performance monitoring data, and having reliable annual budgets. Staff receive training regularly and receive mentoring opportunities.
Evaluation	No evaluations in the last 3 years	Less than one evaluation each year for the last three years	One evaluation per year for the last three years and consistent evaluation use.	One or more evaluation per year for the last three years, consistent evaluation use, and broad communication of results.	One or more evaluation for the last three years, consistent evaluation use, broad communication of results, and a learning agenda.
Evaluation staffing, training and funding	<b>No staff</b> with evaluation functions and <b>no budget</b>	One or two staff members planning for evaluation.	One or two staff members planning for evaluation.	One or two staff members planning for evaluation.	<b>Dedicated full-time staff for evaluation.</b> There are written and utilized evaluation position descriptions



Domain	Not Performing	Evolving	Performing	High Performing	Very High Performing
	for evaluation projects.	Staff <b>are developing</b> budgeting, record keeping, and management systems for evaluation.	Staff <b>have established</b> budgeting, record keeping, and management systems for evaluation.	Staff <b>have established</b> budgeting, record keeping, and management systems for evaluation. <b>Staff have access to training and take it.</b>	There are established budgeting, record keeping and management systems for evaluation. <b>Staff take advantage of training and mentoring opportunities.</b>
Statistics	Direct training for enterprise data skills does not exist. There is no standard approach to capture skill needs, and data-related positions are defined at the project level.	Training is driven by need. There is no shared understanding of enterprise data in the organization. Skills are assessed on an as-needed basis at the bureau or office level. External support is needed for data related skills.	Enterprise awareness exists for the data skills needed and training is offered to some bureaus or offices. A resource strategy is established to meet needs. Internal data experts are identified that help mentor others.	Common knowledge and requirements for data management skills are shared across the enterprise and training of staff is a standard process. Staffing needs are inventoried and managed by an enterprise-wide governing body.	Enterprise wide knowledge of data exists at all levels and knowledge sharing is encouraged. Training programs are reviewed and optimized to meet needs. Data drives decision-making and business strategy. Needed skills are continuously assessed.
Research and Analysis	Proposals to introduce new or reform existing policy, programs, or activities do not yet rely on systematically collected data and analysis.	Policy, program or activity proposals are based on goals but supporting data relies on limited solutions or options.	Rigorous and well documented data is used to develop policy, program, or activity solutions, but with some significant gaps in evidence and analysis.	Rigorous and well documented research and data are used to develop policy, program, or activity solutions, gaps in evidence are acknowledged, and plans to improve are in place.	Rigorous and well documented research and data is used to develop policy, program, and activity solutions, and gaps in evidence are addressed. Long- and short-term research projects build a foundation of information that can be used as the need arises. Standard procedures are utilized to disseminate research and learning across the bureau and inform decision-making.





## APPENDIX E: CAPACITY ASSESSMENT SURVEY

The Bureau of Budget and Planning (BP) and the Office of Foreign Assistance (F) are collecting information on the Department of State's ability to generate and use evidence in fulfillment of the Foundations for Evidence-Based Policymaking Act (Evidence Act).

This survey's purpose is to assess the Department's capacity in evidence-building activities related to performance monitoring, evaluation, statistics, research and analysis. F and BP will use the information from this survey and other work to plan capacity-building efforts and develop a baseline against which the Department will continue to assess its progress.

The aggregated agency-level information will be shared with the Office of Management and Budget (OMB) as part of the draft Joint Strategic Plan per the Evidence Act, but not individual bureau-level information. Bureau-level responses will be **confidential** – only the bureau or independent office, and the F and BP Evidence Team staff will have access to the data.

We are asking the senior expert in evidence-building activities to lead completion of this survey. This would be a manager most responsible for performance monitoring, evaluation, statistics, research and analysis. **The responses should reflect the perspective of the whole bureau (for example, AF or DS) or independent office (for example, S/GWI or OFM). Please do not assess individual-level capacity, nor the capacity of your immediate team or office.**

### Demographics

To start, we would like to learn about your bureau or independent office and your role. Please note that you should complete this survey on behalf of your bureau (for example, AF or DS) or independent office (for example, S/GWI or OFM). Please do not assess individual-level capacity, nor the capacity of your immediate team or office.

1. Please select your bureau or independent office from the list (Select one from drop-down in SurveyMonkey). *We are asking for this information to compute survey completion rates.*
2. What is your job title? \_\_\_\_\_
3. How many years have you worked within your bureau / independent office, regardless of status (e.g., contractor, direct hire)? If between years (3 years, 10 months), please round down. (Select one from drop-down or radio button)
  - Less than one year
  - 1-3
  - 4-6
  - 7-10
  - 11-15
  - 16 - 20
  - 21 or more



4. How many years have you worked at the Department of State, regardless of status (e.g., contractor, direct hire)? If between years (3 years, 10 months), please round down (3 years). (Select one from drop-down or radio button)
  - Less than one year
  - 1-3
  - 4-6
  - 7-10
  - 11-15
  - 16 - 20
  - 21 or more
  
5. Do you currently serve as a Bureau Evaluation Coordinator (BEC)? (Select one)
  - Yes
  - No
  - Not sure
  
6. Do you currently serve as a Bureau Planner? (Select one)
  - Yes
  - No
  - Not sure

### Definitions

*As you complete the survey, you will notice a few terms. For the purpose of this survey, the following terms are defined as follows:*

- **Evidence-building activities** are any activity related to evaluation, performance monitoring, statistics, research and analysis (e.g., data analysis, policy analysis). This relates to planning, conducting, and commissioning these activities.
- **Performance monitoring** is an ongoing system of gathering information and tracking performance to assess progress against established goals and objectives.
- **Evaluation** is the systematic collection and analysis of information about the characteristics and outcomes of programs, projects, or processes. Evaluation is distinct from assessment which may be designed to examine country or sector context to inform program or project design.
- **Statistics** is collecting, compiling, processing, or analyzing data for the purpose of describing or making estimates about the whole versus the individual. Statistical analysis provides information and evidence on economic, demographic, business, and other trends, and allows basic research to explore theories and test new ideas and helps validate hypotheses about performance as compared to desired results.
- **Research and analysis** is any non-evaluation activity which is a rigorous study directed at understanding a subject, applying new knowledge to meet a recognized need, or applying that knowledge toward the production of useful materials, systems, or methods to meet specific requirements. This would not include short-term analysis to manage annual programs, but rather longer-term exploration of policies and programs through literature reviews, academic studies, field work, and other methods. It can be conducted by the Department or on behalf of the Department by universities, NGOs, or other outside sources. This does not include routine performance monitoring.

### Survey



To start, we would like to learn more about the evidence-building activities (performance monitoring, evaluation, statistics, research and analysis) at the Department and who is conducting them.

7. Please provide the office name and symbol (e.g., F/Planning, Performance, and Systems (F/PPS); DS/Office of Management Services/Policy and Planning Division (DS/MGT/PPD)) of any division or office **with core functions in or dedicated** to evidence-building activities. For your bureau or independent office, please enter as many as needed.

*Core responsibilities are those defined in job descriptions and formal responsibilities. Please do not include program officers who conduct routine performance monitoring as part of their project oversight duties.*

Write-in:

8. Over the last three years, evidence-building activities have been carried out by... (please select all that apply):

- Federal employees
- Onsite contractors
- Implementing partners (e.g., grantees, interagency partners, contract firms)
- Academic partners
- Other: [Write-in]

9. For your bureau or independent office, please review the list of evaluations produced in the past 3 years, as listed in the Evaluation Registry (ER) or Evaluation Management System (EMS). If any are missing, please use the links below to add any other evaluations that the bureau may have conducted in that time frame. *[Please check pre-populated data in SurveyMonkey]*

- Please input foreign assistance evaluations in the Evaluation Registry: <http://nextgen.dfafacts.gov/> (accessible on GO Browser or Virtual)
- Please input diplomatic engagement evaluations in the Evaluation Management System: <http://pps.bp.state.sbu> (accessible on GO Virtual)

10. For your bureau or independent office, how many full-time staff members--regardless of hiring mechanism (e.g., onsite contractor, direct hire)--have **core responsibilities** related to evaluation, statistics, or research and analysis.

*Core responsibilities are those defined in job descriptions and formal responsibilities. Please do not include program officers who conduct routine performance monitoring as part of their project oversight duties. [If zero (0), please skip to Question 11.]*

Amount of time spent on evidence-building activities	Number of individuals
Full-time (at least 75% of time)	
Half-time (approximately 50-75% of time)	



Partial (approximately 25-50% of time)	
Minimal (less than 25% of time)	

11. For this item, we are interested in understanding how staff spend their time on evidence-building activities. On **average**, how do the staff included in **Question 10** spend their time on evidence-building activities? *Please input percentages or zero for all categories.*

	Full-time (at least 75% of time)	Half-time (approximately 50-75% of time)	Partial (approximately 25-50% of time)	Minimal (less than 25% of time)
<b>Performance monitoring</b>	%	%	%	%
<b>Evaluation</b>	%	%	%	%
<b>Statistics</b>	%	%	%	%
<b>Research and analysis</b>	%	%	%	%

12. Over the past three years, what types of evidence-building activities were have been **consistently carried out** across your entire bureau or independent office?

Please select those activities that are most **consistently conducted** by staff (i.e., do not select items that are not systematically part of your processes or are simply a best practice of one or a few staff members). *[Please select all that apply.]*

- Collecting performance monitoring data/project indicator data
- Managing performance monitoring data/project indicator data, including data input, within a central repository or database
- Analyzing performance monitoring data/project indicator data
- Reviewing or assessing milestones and performance indicators related to bureau strategic plan goals and objectives
- Conducting periodic data quality checks
- Conducting site visits/activity monitoring trips
- Holding periodic check-in meetings (monthly, quarterly) with implementing partners or contractors
- Reviewing quarterly or other periodic reporting for funded projects
- Compiling project achievements/outcomes (narrative data) in a central repository or database
- Commissioning external evaluations of your programs
- Commissioning external research and analysis activities
- Conducting internal research and analysis activities (conducted by internal staff)
- Conducting internal evaluation (conducted by internal staff)

13. Thinking about the data collected in support of evidence-building activities, with which of the following audiences are these findings **most often** shared? *Please select all that apply:*

- Bureau program officers
- Bureau leadership (e.g., Deputy Directors, Directors)
- Bureau Front Office (e.g., DAS, PDAS, A/S)
- Other Department of State bureaus or independent offices
- Other federal agencies
- Congress
- Academic researchers



- Implementing partners (e.g., grantees)
- The public
- We do not routinely share findings
- Other: [fill-in]

14. Does your bureau have and maintain a learning agenda (or evaluation/research plan)?

*A **learning agenda** or evaluation plan is a plan for identifying and addressing questions relevant to your bureau or office's programs, policies, and regulations. It may be comprise of a list of priority question that, when answered by learning activities such as research and evaluation. **A listing of existing evaluation documents does not meet this criteria.** [Select one]*

- Yes
- No
- Not sure

Please enter any comments here, if not sure:

15. Thinking about how **program design and performance monitoring** is done in your bureau or independent office, which description best fits your practice at this time? *Performance monitoring is defined as an ongoing system of gathering information and tracking performance to assess progress against established goals and objectives.*

Program(s) have not yet been designed and performance monitoring systems have not yet been established.	At least one program has been designed (i.e., logic model or similar framework developed) and monitoring activity is occurring.	Some programs are designed, monitoring plans align to logic models and regular monitoring activities occur.	Many programs are designed, monitoring plans align to logic models, regular monitoring activities occur.  Processes for resolving program issues are monitored. Evaluation recommendations are developed.	Nearly all programs are designed, monitoring plans align to logic models, regular monitoring activities are occurring. Processes for resolving program issues are monitored. Staff factor in data, evaluation recommendations, and return on investment in programming decisions.
---	---	---	---	---

16. Thinking about how **program design and performance monitoring** are **staffed** and **training** provided in your bureau or independent office, which description best fits your practice at this time? (Select one)



No available staff to manage <b>program design</b> and performance monitoring	External subject matter experts are engaged to develop program design and performance monitoring because it is not a responsibility for internal staff.	Staff responsibilities include designing programs, and collecting performance monitoring data.	Staff responsibilities include designing programs, collecting performance monitoring data, and having annual funding to improve processes. Staff are periodically receiving training.	Staff responsibilities include designing programs, collecting performance monitoring data, and having reliable annual budgets. Staff receive training regularly and receive mentoring opportunities.
---	---	--	---	--

17. Thinking about how **evaluation** is **staffed** and **planned** in your bureau or independent office, which description best fits your practice at this time? (Select one)

<b>No staff</b> with evaluation functions and <b>no budget</b> for evaluation projects.	One or two staff members planning for evaluation.  Staff <b>are developing</b> budgeting, record keeping, and management systems for evaluation.	One or two staff members planning for evaluation.  Staff <b>have established</b> budgeting, record keeping, and management systems for evaluation.	One or two staff members planning for evaluation.  Staff <b>have established</b> budgeting, record keeping, and management systems for evaluation. <b>Staff have access to training and take it.</b>	<b>Dedicated full-time staff for evaluation.</b> There are written and utilized evaluation position descriptions  There are established budgeting, record keeping and management systems for evaluation. <b>Staff take advantage of training and mentoring opportunities.</b>
---	--	--	--	---

In this section, we are exploring how evidence-building activities are utilized and applied. As an expert, please answer these items to the best of your ability.

18. In **Question 12**, you reported consistently conducting the following evidence-building activities. Please rate how useful the following pieces of evidence are in helping staff make programmatic decisions (e.g., informing project design, planning, and/or budgeting; making recommendations to stop, start, or keep certain approaches). [Categories piped from Question 12]

	Not useful	Slightly useful	Somewhat useful	Moderately useful	Extremely useful
Collecting performance	[]	[]	[]	[]	[]





monitoring data / project indicator data					
Managing performance monitoring data / project indicator data into a central repository / database					
Analyzing performance monitoring data / project indicator data					
Conducting periodic data quality checks					
Conducting site visits for funded projects					
Holding periodic check-in meetings (monthly, quarterly) with implementing partners					
Reviewing quarterly reporting for funded projects					
Compiling project achievements / outcomes (narrative data) in a central repository / database					
External research and analysis activities outside the PMES IDIQ					
Conducting internal					



assessments, analysis, studies (conducted by Bureau staff)					
External evaluations (IDIQ or non-IDIQ funded)					

19. In your bureau or independent office, do you agree that those responsible for designing and managing programs have the time to use evidence to inform project design or make adjustments to current operations. [Please select the best option]

1: Strongly disagree	2: Disagree	3: Neither disagree nor agree	4: Agree	5: Strongly agree
----------------------	-------------	-------------------------------	----------	-------------------

20. In the table below, please rate the quality of the following pieces of evidence that your bureau or independent office collects. [Categories piped from Question 12]

	1: Poor. Not systematically planned or conducted.	2: Fair. Fragmented planning and operationalization.	3: Good. Sufficient quality.	4: Very good. High quality evidence with limitations but no serious flaws..	5: Excellent Fully reflects best practices.
Collecting performance monitoring data / project indicator data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Managing performance monitoring data / project indicator data into a central repository / database	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Analyzing performance monitoring data / project indicator data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Conducting periodic data quality checks	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Conducting site visits for	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



funded projects					
Holding periodic check-in meetings (monthly, quarterly) with implementing partners	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Reviewing quarterly reporting for funded projects	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Compiling project achievements / outcomes (narrative data) in a central repository / database	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
External research and analysis activities outside the PMES IDIQ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Conducting internal assessments, analysis, studies (conducted by Bureau staff)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
External evaluations (IDIQ or non-IDIQ funded)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Next, we would like to learn more about the types of methods that are used and their application to evidence-building activities (performance monitoring, evaluation, research and analysis, and statistics).**

21. From **Question 12**, you reported that your bureau or independent office consistently conducts the following evidence gathering activities. For each evidence type, please answer the following question:  
*The appropriate combination of methods is used to address our operational and learning needs:*

*Methods are appropriate when the processes used meet the objectives for the task.*



	1: Strongly disagree	2: Disagree	3: Neither agree nor disagree	4: Agree	5: Strongly agree
Collecting performance monitoring data / project indicator data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Managing performance monitoring data / project indicator data into a central repository / database	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Analyzing performance monitoring data / project indicator data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Conducting periodic data quality checks	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Conducting site visits for funded projects	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Holding periodic check-in meetings (monthly, quarterly) with implementing partners	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Reviewing quarterly reporting for funded projects	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Compiling project achievements / outcomes (narrative data) in a central repository / database	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
External research and analysis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



activities outside the PMES IDIQ					
Conducting internal assessments, analysis, studies (conducted by Bureau staff)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
External evaluations (IDIQ or non-IDIQ funded)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

22. Thinking about how your bureau or independent office uses research and analysis, which description best fits your practice at this time?

***Research and analysis** is any non-evaluation activity which is a rigorous study directed at understanding a subject, applying new knowledge to meet a recognized need, or applying that knowledge toward the production of useful materials, systems, or methods to meet specific requirements. This would not include short-term analysis to manage annual programs, but rather longer-term exploration of policies and programs through literature reviews, academic studies, field work, and other methods. It can be conducted by the Department or on behalf of the Department by universities, NGOs, or other outside sources. This does not include routine performance monitoring.*

Proposals to introduce new or reform existing policy, programs, or activities do not yet rely on systematically collected data and analysis.	Policy, program or activity proposals are based on goals but supporting data relies on limited solutions or options.	Rigorous and well documented data is used to develop policy, program, or activity solutions, but with some significant gaps in evidence and analysis.	Rigorous and well documented research and data are used to develop policy, program, or activity solutions, gaps in evidence are acknowledged, and plans to improve are in place.	Rigorous and well documented research and data is used to develop policy, program, and activity solutions, and gaps in evidence are addressed. Long- and short-term research projects build a foundation of information that can be used as the need arises. Standard procedures are utilized to disseminate research and learning across the bureau and inform decision-making.
--	--	---	--	--



23. What are the greatest barriers exist for your bureau or independent office when it comes to using rigorous methods for evidence-building activities, including performance monitoring, evaluation, statistics, research and evaluation? *[Please select up to five].*

*Rigorous is defined as systematic and explicit/transparent procedures that are appropriate for the type of method used (e.g., qualitative, experimental, data management).*

- Insufficient time for staff to design external evidence-building activities (e.g., writing or refining evaluation contract statements of work)
- Insufficient time for staff to conduct evidence-building activities
- Insufficient time for staff to manage external evidence-building activities (e.g., external evaluations, external research)
- Insufficient number of staff focused on evidence-building activities
- Lack of appropriate skills to conduct evidence-building activities
- Lack of appropriate knowledge to manage evidence-building activities
- Lack of funding for evidence-building activities (e.g., external evaluations)
- Lack of demand from stakeholders
- No applicable barriers
- Other (please explain)

**The following questions address managing partiality and promoting independence in performance monitoring, evaluation, statistics, research and analysis activities.**

24. For each evidence type, please rate your level of agreement with the following statement: Our bureau is able to **mitigate inappropriate influence** in **internal** evidence-building activities.

*Definition: **Inappropriate influence** are mitigated when evidence from activities (e.g., performance monitoring, evaluation, research and analysis, statistics) are systematically and fairly considered regardless of the findings. **Internal** activities are those conducted by staff in the bureau / independent office, regardless of hiring mechanism (e.g., direct hire, onsite contractors).*

	1: Strongly disagree	2: Disagree	3: Neither agree nor disagree	4: Agree	5: Strongly agree	Not sure
Performance monitoring	[]	[]	[]	[]	[]	[]
Evaluation	[]	[]	[]	[]	[]	[]
Statistics	[]	[]	[]	[]	[]	[]
Research and analysis	[]	[]	[]	[]	[]	[]

25. For each evidence type, please rate your level of agreement with the following statement: Our bureau is able to mitigate inappropriate influence in **external** evidence-building activities.

*Definition: **Inappropriate influence** is mitigated when evidence from activities (e.g., performance monitoring, evaluation, research and analysis, statistics) are systematically and fairly considered regardless of the findings. **External** activities are those conducted by individuals outside the bureau or independent office (e.g., external consultants, offsite contractors, implementing partners).*





	1: Strongly disagree	2: Disagree	3: Neither agree nor disagree	4: Agree	5: Strongly agree	Not sure
Performance monitoring	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Evaluation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Statistics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Research and analysis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

In the following section, we are interested in learning more about the types of skills present within bureaus and independent offices, as well opportunities for greater capacity building.

26. For each method, please rate your level of agreement with the following statement: Our bureau or independent office has staff who have the skills to use the following methods to generate evidence:

	This skill is not needed in our bureau or independent office	1: Strongly disagree	2: Disagree	3: Neither agree nor disagree	4: Agree	5: Strongly agree
Fundamental statistics (descriptive, sample design)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Advanced statistics (regression, inferential, Bayesian)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Experimental or quasi-experimental design	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fundamental qualitative collection methods (interview design, focus groups)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Qualitative data analysis (content analysis; thematic analysis; computer-assisted, e.g., NVivo)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Various qualitative research/evaluation designs (narrative inquiry, Most Significant Change, process tracing)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



Data management (structuring data in Excel or other repositories)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Data analytics (Excel, R, SPSS, Power BI, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Data visualization (Tableau, Power BI, R, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

27. Our bureau or independent office supports training for direct hires interested in evidence-building activities. [If 'Yes', skip to Question 29),

- Yes
- No
- Not sure

28. [If the answer to Question 27 is 'No' or 'Not Sure'] What is the largest obstacle to such training?

- Funding
- Time
- No applicable training
- Not approved by supervisors
- Other (please specify): \_\_\_\_\_

29. Of the following Department resources, which do you consider most useful for your evidence-building activities. [Please rank answer choices in order of importance or mark as N/A]

- Training (Managing Evaluations course, Strategic Planning and Performance Management (PA 315))
- Policies and processes (18 FAM 300 covering strategic planning, performance management, and evaluation)
- Online resources (Managing for Results website, Evaluation IDIQ contract vehicle website, Program Design and Performance Management toolkit)
- Communities of practice (Evaluation community of practice, Program Design and Performance Management COP, Bureau Planners group)

30. What type of information or tools could help you improve evidence-building activities? Please select all that apply.

- Additional policy guidance (e.g., 18 FAM 300)
- Short guidance notes (e.g., step-by-step instructions, how-to notes) on evidence generation, management, dissemination, and use
- Additional guidance on commissioning and procuring evidence-building activities
- Direct training/technical assistance from a Department evaluation advisor
- Direct training/technical assistance from external experts (e.g., academics, researchers)
- Tools for data management



- Tools for data analysis
- Tools for data visualization
- Tools for sharing and disseminating evidence
- Access to data from within the Department
- Access to data from external sources
- Access to an advisor who can assess methodological or analytical questions
- Additional funding for research, M&E, and learning
- Other – Please specify:
- None – Our bureau/independent office does not need additional resources
- I don't know

**Additional Comments**

31. If you have any additional thoughts or questions about the Department's performance monitoring, evaluation, research and analysis and statistics, please use the box below to share. [Open-ended]

32. This has been cleared by Bureau leadership. Please include name here:

