INSTITUTE FOR DEFENSE ANALYSES

CLEARED
For Open Publication

Jan 27, 2020

Summary of DoD STEM Workshop on Evaluation

James Belanich
Daniel Kolodrubetz
Emily A. Fedele
Christian Dobbins
Franklin L. Moses

About This Publication

This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-14-D-0001, Project AI-2-4016, “STEM (Science, Technology, Engineering and Mathematics) Program Evaluation,” for the STEM Development Office in the Office of the Assistant Secretary of Defense (Research and Engineering) Research Directorate. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

For More Information

James Belanich, Project Leader
jbelanic@ida.org, 703-845-6606

Leonard J. Buckley, Director, Science and Technology Division
lbuckley@ida.org, 703-578-2800

Copyright Notice

© 2019 Institute for Defense Analyses
4850 Mark Center Drive, Alexandria, Virginia 22311-1882 • (703) 845-2000.

This material may be reproduced by or for the U.S. Government pursuant to the copyright license under the clause at DFARS 252.227-7013 (Feb. 2014).
Summary of DoD STEM Workshop on Evaluation

James Belanich
Daniel Kolodrubetz
Emily A. Fedele
Christian Dobbins
Franklin L. Moses
Executive Summary

Background

Within the Department of Defense (DoD), many agencies and organizations participate in science, technology, engineering, and mathematics (STEM) outreach and education efforts. These efforts align with the goals laid out in Federal and DoD STEM strategic plans and are coordinated across Components by the DoD STEM Office. A key part of the strategic plan is transparency and accountability, so evaluating the various efforts is essential. This report discusses a workshop that was held to bring together participants from multiple DoD Components, Federal agencies, academic institutions, and other organizations to provide a high-level overview of STEM priorities across the DoD and address effective program evaluation.

Summary of Workshop

The workshop was split into two main sections. In the morning speakers provided bigger picture goals for STEM in general and STEM evaluation specifically. The afternoon focused on examples of evaluation best practices. In between, a breakout group session discussed participant roles and experiences with program evaluations.

Morning Session

The morning presentations emphasized the importance of evaluations for STEM success, which was the topic of the keynote address presented by Dr. Jihfen Lei, the Deputy Director for Research, Technology and Laboratories in the Office of the Under Secretary of Defense for Research and Engineering. Her talk was followed by Dr. Jon Werner-Allen’s description of how the new Federal STEM Strategic Plan provides a path to successful evaluations by advocating transparency and accountability through four objectives:

- Leveraging and scaling evidence-based practices across STEM communities.
- Reporting participation rates of underrepresented groups.
- Using common metrics to measure progress.
- Making program performance and outcomes publicly available.

The OSD STEM office perspective, presented by Mr. Louie Lopez, emphasized two reasons for the importance of evaluations: to know that what we are doing works and to
measure our impact as a means of telling our story. The introductory presentations were concluded by highlighting a key theme of the strategic plan: the need to better understand underrepresented groups in STEM. Dr. Diann McCants discussed diversity and inclusion considerations in evaluations, making two key points: (1) a diverse STEM talent pool leads to a better organization, and (2) culturally sensitive evaluations can help build diversity by giving a better understanding of the real landscape of STEM education.

A panel of STEM professionals from outside DoD served to complement DoD’s perspective. The panelists included Dr. Sarah-Kay McDonald, National Science Foundation; Dr. John Baek, National Oceanic and Atmospheric Administration; Mr. Rick Gilmore, National Aeronautics and Space Administration; and Dr. Carol O’Donnell, Smithsonian Science Education Center.

Their discussion, including questions from attendees, broached a broad range of topics and had three major takeaways:

- Examine programs carefully to determine the appropriate level of evaluation.
- Demographic/personal data can be hard to collect, but they are important to show impact of the program.
- Resources are a concern for evaluations, but there are cost-cutting techniques.

**Breakout Group**

After the panel, workshop attendees were assigned to breakout groups to discuss the evaluation challenges, solutions, benefits, and accomplishments that they have seen. The table below summarizes the main takeaways in each of those areas.

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Solutions</th>
<th>Benefits/Accomplishments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finding resources</td>
<td>More strategic evaluation planning</td>
<td>Iterative program improvement and maximized outcomes</td>
</tr>
<tr>
<td>Motivation for data collection</td>
<td>Establishing evaluation context</td>
<td>Strategic support</td>
</tr>
<tr>
<td>Standardized data collection</td>
<td>Defining common data metrics</td>
<td>Sharing of evaluation tools and findings</td>
</tr>
<tr>
<td>Acquiring analyzable data</td>
<td></td>
<td>Transparency</td>
</tr>
</tbody>
</table>

**Afternoon Session**

Four diverse examples of evaluation best practices highlighted the afternoon presentations. An overview of the STARBASE evaluation, presented by Mr. Gary Behrens, emphasized the need to plan as much as possible, communicate often, and keep innovating.
The Army Education Outreach Program (AEOP) portfolio, presented by Dr. Carla Johnson, provided insight into the challenges and possibilities when evaluating a large scale portfolio. The role of educational principles in STEM evaluation was the perspective advocated by Dr. Adam Maltese. He provided empirical evidence of the importance of frequent iterative cycles of both quantitative and qualitative evaluations, getting multiple perspectives, and using the STEM participants’ views. In the final presentation of the day, Dr. Asha Balakrishnan discussed an evaluation of the Science, Mathematics, and Research for Transformation (SMART) scholarship, which feeds directly into the DoD workforce. Each of these four STEM programs is successful and provides evaluation guidance that others could follow.

Conclusions and Next Steps

The workshop organizers followed up with the participants, sharing everyone’s contact information to begin building a community of practice, and providing presentations used by the workshop’s speakers. A community of practice is a way going forward to keep the lines of communication open among workshop attendees to provide a source of ideas for STEM practitioners. Information and resources from the workshop will form the beginning of a repository of evaluation best practices.
## Contents

1. Introduction .................................................................................................................1  
   A. Background: DoD as a STEM Organization..........................................................1  
   B. Workshop Structure.......................................................................................3  
2. Workshop Proceedings (Presentations and Group Discussion) ............................5  
   A. Keynote Address: The Importance of Evaluations for STEM Success.............5  
   B. Federal STEM context and CoSTEM Strategic Plan.......................................7  
   C. OSD STEM Perspective..................................................................................8  
   D. Diversity and Inclusion Considerations in Evaluations ..................................8  
   E. Panel: STEM Evaluations Outside DoD .........................................................9  
      1. The Panelists and Agencies Introduction ..............................................10  
      2. Evaluation Planning .............................................................................11  
      3. Collecting Demographic Data ................................................................12  
      4. Cost Concerns ......................................................................................13  
   F. Breakout Groups: Discuss Evaluation Issues.................................................13  
      1. What challenges have you seen in your program evaluations? ...............14  
      2. What solutions have you implemented for those challenges? ..............15  
      3. What are the benefits you see from performing evaluations? ...............16  
      4. What are some of your organizations’ accomplishments in STEM evaluation?..................................................................................................17  
   G. Program Evaluation Example: STARBASE ..................................................17  
      1. An Overview of STARBASE...................................................................18  
      2. The Evaluation ......................................................................................18  
      3. Lessons Learned....................................................................................19  
   H. Portfolio Evaluation Example: Army Education Outreach Program .............19  
      1. Overview ...............................................................................................19  
      2. Evaluation ............................................................................................19  
      3. Lessons Learned....................................................................................20  
   I. Principle Evaluation Example: Pipeline Persistence ......................................21  
      1. Overview ...............................................................................................21  
      2. Evaluation ............................................................................................21  
      3. Lessons Learned....................................................................................22  
   J. STEM Workforce Program Evaluation Example: SMART Scholarship ........22  
      1. Overview ...............................................................................................22  
      2. Evaluation ............................................................................................23  
      3. Lessons Learned....................................................................................25  
   K. Closing and Next Steps .....................................................................................25  
3. Summary..................................................................................................................27
Appendix A. Planned Workshop Agenda ................................................................. A-1
Appendix B. List of Participants ........................................................................... B-1
Appendix C. Diversity and Inclusion in STEM Evaluation:
   Quick Knowledge Check .................................................................................... C-1
Abbreviations ........................................................................................................ D-1
1. Introduction

The Department of Defense (DoD) employs a large scientific and technologically skilled workforce and therefore has a vested interest in supporting science, technology, engineering, and mathematics (STEM) education at all levels across the nation. In the future, students who acquire STEM skills may become part of the talent pool that feeds the highly qualified DoD workforce. The mission of the DoD STEM enterprise is “to attract, inspire, and develop exceptional STEM talent across the education continuum and advance the current STEM workforce to meet current and future defense technological challenges.” To accomplish that mission, DoD STEM has five major portfolio goals: (1) build a high-qualified STEM workforce overall and in mission-critical areas; (2) enhance the preparation of dependents of members of our Armed Forces for careers in STEM; (3) provide education and outreach programs and activities that build the pipeline; (4) promote increased participation of underserved groups; and (5) communicate the value of STEM investments as a critical enabler to the DoD mission.

This report is about the DoD STEM Evaluation workshop, which was held to help facilitate the goals of the DoD STEM enterprise. The workshop for STEM professionals provided a high-level overview of STEM priorities across DoD and the Federal government with a focus on program evaluations. The workshop brought together participants from DoD agencies, Federal agencies, academic institutions, and a few organizations that provide support to DoD STEM efforts. This report describes the proceedings of the workshop and gives key takeaways from each session.

A. Background: DoD as a STEM Organization

Within the DoD many agencies and organizations need continued access to a STEM-skilled pool of talent for the future workforce. Most of these organizations also have a current STEM-skilled workforce that may participate in STEM education and outreach efforts to inspire and develop our Nation’s students, who may join them in the future. This STEM-skilled workforce resides at the numerous Service laboratories and engineering centers along with other agencies and organizations that have technological mission requirements. Because skills and STEM requirements are dispersed across DoD’s agencies and organizations, the DoD STEM Office functions in a coordination, policy, and oversight role of Defense-wide STEM education, outreach, workforce development programs and initiatives, as well as the execution of joint Service and Agency activities.
The broad geographic range covered by the Services and other DoD organizations can be considered a strength of DoD STEM educational capabilities because the target audience (i.e., PK-12 students, graduate and undergraduate students, STEM teachers, and parents) are located across the country. Also, because the DoD STEM workforce covers many different academic disciplines and technology fields (e.g., physics, chemistry, biology, psychology, engineering, robotics, computer science, cybersecurity), DoD has the ability to engage and inform students across many STEM fields. At the same time, however, the distribution of DoD STEM educational programs across agencies and locations means that they do not regularly interact with one another to learn and share perspectives.

A workshop was identified as a method for bringing together people involved with DoD STEM educational efforts at multiple levels, such as front-line STEM program providers working directly with students, STEM program managers who may manage multiple programs or a portfolio, and agency leads who may make programmatic decisions about how to best allocate resources. The idea of a DoD STEM evaluation workshop was shared with the DoD STEM community and met with a positive response.

The workshop was developed specifically to promote quality evaluations across DoD in support of the FY2019 Administration Research and Development Budget Priorities. Those priorities highlight the Nation’s need to develop a future-focused STEM workforce through education programs along with a means to evaluate those programs:

Agencies should give priority to policies and actions that place an emphasis on expanding the STEM workforce to include all Americans, both urban and rural, and including women and other underrepresented groups in STEM fields. In order to track improvements in these areas, agencies should develop quantitative methods or metrics and collect data to analyze the effectiveness of the STEM programs.

To address these priorities, the DoD STEM Office is emphasizing program evaluation and promoting an evaluation and assessment capability.

The DoD STEM Evaluation and Assessment Capability is organized into three levels of analysis: (1) portfolio analysis, (2) program evaluation, and (3) assessment of principles and methods for effective STEM programs. Portfolio analysis addresses a set of programs and efforts with a shared management/funding structure or common goals. This level of analysis may produce data that support decisions in the planning, programing, budgeting, and execution process as an organization seeks to align its set of programs to a higher level strategic plan. Program evaluations take place at the individual program level: evaluating

---

how well a specific program meets its particular goals and objectives. An assessment of principles (e.g., learner development/progression over time) and methods (e.g., hands-on activities) addresses when and how particular principles or methods may be applicable across programs. This allows lessons learned from one program to be available for use by other programs. The importance of STEM program evaluation and related levels of analyses informed the planning for the workshop described in this report.

B. Workshop Structure

Based on discussions with the DoD STEM Office and STEM program managers, we decided the workshop would have two main goals:

1. Build a community of practice to promote communication about evaluations across DoD STEM.

2. Expose participants to a range of high-quality evaluations from DoD STEM and the Federal government in an effort to help organizations and participants develop and conduct the appropriate level of evaluation for DoD STEM programs that may inform data-driven decisions.

The workshop began with a high-level overview of STEM priorities across DoD and the Federal government to provide the participants with perspective on the importance of STEM educational programs. This broad perspective was then supplemented with information from a range of other government organizations highly interested in STEM educational programs. After this initial set of presentations, an active component (i.e., breakout group session) allowed all participants to engage in discussion about STEM evaluations and meet other professionals in the field from different agencies. Then, four examples of high-quality evaluations—a portfolio analysis, two program evaluations (one focused on teaching children and the other on workforce development), and an assessment of principles and methods—were presented to highlight evaluation capabilities and emphasize the value in evaluation. Appendix A gives the full agenda for the workshop.
2. Workshop Proceedings
(Presentations and Group Discussion)

This section summarizes each of the sessions during the workshop, highlighting key details and important takeaways from each session.

A. Keynote Address: The Importance of Evaluations for STEM Success

Dr. JihFen Lei, the Deputy Director for Research, Technology and Laboratories in the Office of the Under Secretary of Defense for Research and Engineering, provided a high-level DoD perspective on STEM education programs and evaluation. Dr. Lei is responsible for establishing and implementing the DoD S&T strategic vision, as well as identifying and influencing the DoD-wide planning guidance. Her presentation covered three themes: (1) STEM pipeline and workforce development, (2) STEM domains of priority to DoD, and (3) evaluation to provide quality information to decision-makers.

Dr. Lei described how it is important for DoD to work together, continuing to foster the STEM pipeline and enhancing the preparedness of our future workforce. The link between STEM programs and workforce was highlighted in the context of DoD’s recent reorganization of the research and development structure. What was previously the Acquisition, Technology, and Logistics (AT&L) Office is now divided into two offices: the Office of the Under Secretary of Defense for Research and Engineering (OUSD (R&E)) and the Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD (A&S)). The creation of OUSD (R&E) recognized the importance of R&E and the Department’s need for an organization skilled in STEM capabilities that is focused on R&E to achieve technological superiority for our military.

Dr. Lei described how the Department is in a great position to make a unique contribution to STEM education by leveraging DoD’s 205 laboratories and research facilities and the 50,000+ scientists and engineers that work in them. One positive outcome of DoD STEM investments might be that the students involved in our STEM programs eventually end up in DoD labs. If they end up in academia or industry, however, it is still a positive outcome for DoD because DoD works closely with academia and industry in research and technology development. Linking the theme of workforce capability and research domains, Dr. Lei indicated that DoD STEM efforts do quite a bit to raise awareness of the DoD science and technology (S&T) enterprise. The connections created
between students participating in scholarship and internship programs, faculty, and DoD lab personnel help foster new partnerships leading to long-term relationships for the DoD.

Dr. Lei then transitioned to a discussion of domains of STEM that have been prioritized by DoD, guided by the National Defense Strategy (NDS)\(^2\) that was rolled out last year. The NDS guides all DoD planning, including the strategy for STEM efforts. Based on the NDS, OSD R&E has identified 10 modernization priorities: microelectronics; cyber; quantum science; directed energy; machine learning/artificial intelligence; fully networked command, control, and communication; space; autonomy; hypersonics; and biotechnology\(^3\). Dr. Lei stated that she believes investments in STEM could support these priorities.

The theme of evaluation was highlighted as a means to provide quality information to decision-makers. Conducting proper evaluations of STEM programs allows the department to make informed decisions to ensure that we continue to invest in programs that will have a positive impact. Dr. Lei outlined the need for evaluations of STEM programs to ask the “so-what” question (i.e., testing whether the program is having an important impact). This is crucial as we plan and develop policy decisions that align with the goals and priorities of the DoD.

Dr. Lei noted that the role of STEM education may not be easily quantifiable, but the innovative discussions at this workshop and follow-up efforts to promote STEM education and outreach will help DoD achieve its goals in new and more effective ways. Dr. Lei emphasized that STEM education programs are supported by Dr. Lisa Porter, the Deputy Under Secretary of Research and Engineering, who has described her core principals as an unwavering commitment to technical excellence and technical truth; the essential value of peer review; the importance of technical integrity; the acceptance of some failure, but not if it is due to ignoring the laws of physics or not providing sufficient programmatic oversight on the execution of your idea; and the importance of planning in the right direction by using a physics analogy of speed versus velocity: velocity is speed in a particular direction, and it is the velocity that is important rather than just speed; going fast does you no good if you are heading off a cliff.

Dr. Lei ended her presentation by encouraging the participants to be a part of the discussions and help DoD continue to expand our STEM horizons: “It is in groups like this where we can come up with the new ideas, and adjust our current thinking, to make progress in such an important area.”

\(^3\) https://www.cto.mil/modernization-priorities/
B. Federal STEM context and CoSTEM Strategic Plan

Dr. Jon Werner-Allen of the Office of Science and Technology Policy (OSTP) presented the Federal perspective of STEM evaluation. Dr. Werner-Allen, an American Association for the Advancement of Science (AAAS) Science and Technology Policy Fellow at OSTP, provided insight about how DoD’s efforts fit into the larger national agenda. At OSTP he was instrumental in developing the Federal STEM Strategic Plan and is currently working on the Federal STEM Implementation Plan, which will describe how the Federal agencies intend to work toward accomplishing their STEM goals outlined in the strategic plan.

Dr. Werner-Allen began by describing the new Federal STEM Strategic Plan, which is the result of collaboration from Committee on Science, Technology, Engineering, and Math Education (CoSTEM) of the National Science and Technology Council (NSTC) and OSTP. Many Federal agencies were involved in its development, and academia, non-profit organizations, and industry also provided input.

This national STEM plan promotes three goals, which cut across agencies and stakeholders:

1. Build strong foundations for STEM literacy.
2. Prepare the STEM workforce for the future.
3. Increase diversity, equity, and inclusion in STEM.

Dr. Werner-Allen pointed out that achieving these goals will require different lines of effort, which have direct tie-in with the role of the workshop’s attendees: developing strategic partnerships, engaging students where disciplines converge, building computational literacy, and improving transparency and accountability. This last line of effort is closest to the workshop theme of evaluation.

The Federal STEM Strategic Plan describes four objectives for transparency and accountability, as well as recommended actions. The first of these is leveraging and scaling evidence-based practices across STEM communities, and the Strategic Plan recommends broadly identifying and disseminating effective STEM education programs, practices, and policies, including PK-12, formal and informal, undergrad, grad, and lifelong learning.

The second objective is reporting participation rates of underrepresented groups. Dr. Werner-Allen points out the Army’s Army Education Outreach Program (AEOP) as a good example here. The recommendations are to establish a clear baseline of who is reporting demographic information and of what those participation rates are.

The third objective is to use common metrics to measure progress, which includes developing and reporting additional metrics if needed to assess effectiveness.
The last objective is to make program performance and outcomes publicly available, which can be done by collecting and disseminating templates, examples, best practices, and lessons learned. An online dashboard or similar public resource that shows performance and outcome data for Federal programs could accomplish this.

Dr. Werner-Allen said that actually undertaking these recommendations will be challenging, but having STEM providers, evaluators, and program managers all talk with each other could give some understanding of baseline capabilities in data collection and dissemination and of common challenges, such as defining a participant, defining “rural,” collecting participant (demographic) data, collecting information from different venues and centrally compiling it. There are also legal hurdles involved with data gathered from minors.

C. OSD STEM Perspective

Mr. Louie Lopez, the Director of STEM from the Laboratories and Personnel Office within the Office of the Under Secretary of Defense for Research and Engineering, provided a coordination-of-effort perspective on DoD STEM programs and the value of evaluation. Mr. Lopez explained his perspective on evaluating STEM programs, offering two primary goals in support of evaluation: (1) to know what programs work, and (2) to measure STEM program impact to tell the DoD STEM education story. Because the goals of particular STEM programs vary from one another, Mr. Lopez advocated against counting all STEM engagements in the same way. For example, what happens in a weeks-long Gains in Education of Mathematics and Science (GEMS) or STARBASE activity is different from a 1-day type of engagement like the Science and Engineering Festival, or an 8- to 10-week research apprenticeship at one of our Laboratories. This is a lesson learned from last year’s DoD-wide data collection, and it will be incorporated into the evolution of the DoD portfolio-level evaluation across all Services and Fourth Estate agencies that conduct STEM Education programs. Mr. Lopez explained that over the last couple of years, OSD STEM has looked at improvements in data collection so that it is able to better report our program impact to the Hill and to the Federal STEM Committee, and it will continue with efforts to improve the process.

Mr. Lopez expressed how this workshop is part of the effort to develop a shared vision and mission across DoD agencies. The workshop should help to show what can be achieved together in K-12 STEM efforts, undergraduate and graduate programs, along with workforce initiatives. Mr. Lopez closed with encouragement for the remainder of the day.

D. Diversity and Inclusion Considerations in Evaluations

Dr. Diann McCants of Strategic Analysis supports the OSD Laboratories and Personnel Office. Dovetailing with discussion of the Federal STEM Strategic Plan’s goal of reaching underserved groups, Dr. McCants presented on diversity and inclusion
considerations in evaluations. She began her talk with a quiz probing the audience’s understanding of specific topics in diversity (see the list of true/false questions in Appendix C). These true/false questions highlighted how diversity is a positive characteristic that in some circumstances may be misperceived. The discussion of each item stimulated a dialogue between Dr. McCants and the audience that persisted throughout the talk.

STEM evaluation practices are fundamentally tied to issues of equity because these practices have ingrained values, systems, and behaviors leading to social disadvantage. Focusing on diversity and inclusion in STEM evaluation supports the Federal STEM Education Strategy. Studies have shown that a diverse STEM talent pool leads to stronger diversity of thought and technical ability to be competitive. Therefore, diversity and inclusion matter greatly when trying to attract and develop STEM talent.

Dr. McCants emphasized that representation of underrepresented groups varies by field; better understanding the landscape of STEM education and evaluation will lead to greater equity across the population. Dr. McCants stressed that a more inclusive STEM evaluation process should be culturally relevant and nuanced, reflecting appropriate messaging, which can help eliminate evaluator bias. Having more appropriate messaging will help programs have better buy-in for collecting demographic data and documenting and reporting the participation rates of underrepresented groups. In addition, ensuring the multicultural validity of instruments, measures/metrics, and inferences would better reflect the perspectives of non-dominant groups.

Audience discussion included adjusting messaging with respect to common misconceptions. For example, “increased diversity leads to decreased quality” can be managed with messaging that prioritizes outcomes and achievements. Additional discussion focused on the current makeup of the STEM workforce: diversity of the future STEM workforce depends on increased diversity of the current STEM workforce; the future workforce will suffer if there is no one to look to as an example of diversity. Other strategies to improve diversity include addressing budget decisions and improving data collection.

E. Panel: STEM Evaluations Outside DoD

To provide a broad perspective (i.e., beyond DoD) STEM professionals from around the government presented in a panel format to discuss STEM program evaluation. The panelists spanned a range of roles and agencies:

- **Dr. Sarah-Kay McDonald**, Senior Advisor, National Science Foundation (NSF) Directorate for Education and Human Resources (EHR)
- **Dr. John Baek**, Senior Education Evaluator, National Oceanic and Atmospheric Administration (NOAA) Office of Education
Mr. Rick Gilmore, Evaluation Manager, National Aeronautics and Space Administration (NASA) Office of STEM Engagement

Dr. Carol O’Donnell, Senior Executive, Director, Smithsonian Science Education Center (SEC)

To begin, panelists introduced themselves and their agency’s STEM education and evaluation goals. Then, the discussion turned to three basic questions loosely followed throughout the panel discussion:

- What evaluation methods and techniques have been implemented in your agency?
- What unique evaluation challenges have you run into?
- How are your evaluation results used?

Audience members were encouraged to ask questions during this portion of the panel, and they posed many questions and gave input from their own experiences. The resulting conversation between the panelists and the audience spanned many different topics, summarized in detail below, and there were three main takeaways from the panel discussion:

1. It is important to examine programs carefully to determine the appropriate level of evaluation. This includes talking to all the stakeholders early in the process to get buy-in and feedback.

2. Demographic data and other personal data can be hard to collect, but can be important to show the impact of a program. Developing a solid infrastructure for collecting these data and being culturally responsive to your participants are key for getting reliable responses.

3. Resources are a major concern for performing evaluations, but there are techniques to help bring cost down and to prioritize. Costs can be lessened if an appropriate evaluation strategy is implemented and an appropriate structure is in place for the evaluation to take place.

The remainder of this section focuses on a detailed summary of the panelists’ introductory talking points and the main topics of conversation that emerged from the rest of the discussion.

1. The Panelists and Agencies Introduction

Dr. Sarah-Kay McDonald is a Senior Advisor in the NSF EHR, which has the mission to achieve excellence in U.S. STEM education at all levels in all settings. NSF is focused on discoveries and depends on innovation across sectors, including STEM education. NSF has awards to support professional development to increase the participation of Americans
Dr. Jon Baek is a Senior Education Evaluator for the NOAA Office of Education. NOAA has a diverse set of missions, including engaging the public and strengthening the future workforce. Education and outreach efforts can live under a range of different programs, and there are no appropriated budget lines for evaluation. Dr. Baek tries to take a broad view on evaluation to capture the variety of efforts that NOAA undertakes.

Mr. Rick Gilmore is an Evaluation Manager at the NASA Office of STEM Engagement. He leads a four-person team and distributes evaluation work across the agency. Recently, the NASA Office of STEM Education rebranded to STEM Engagement to reflect more inclusive priorities. The goal is to make powerful connections to NASA’s research. In the evaluation work, Mr. Gilmore uses a learning agenda framework to inform evaluations from start to finish. His long-term goal is to change the culture of evaluation within NASA to make evaluation baked into programs from the beginning of planning.

Dr. Carol O’Donnell is the Director of the Smithsonian SEC, which was founded in 1985. The Smithsonian Institute is a quasi-governmental agency that gets approximately two-thirds of its funding from the Federal government. The museum nature of the Smithsonian poses a challenge for evaluation because it cannot know exactly who walks through the doors. The Smithsonian has split its evaluation goals into three different levels to understand programs with different levels of engagement:

1. Understand whom they reach (e.g., number of website views).
2. Understand whom they engage (e.g., counts at short-term programs).
3. Understand whom they affect (e.g., evaluations of longer term programs).

In an ongoing effort to evaluate programs, the Smithsonian developed a database that is updated with evaluation data monthly and submitted to a publicly available dashboard. The Smithsonian Organization and Audience Research office opened recently to develop methods and techniques for understanding the participants in Smithsonian activities.

2. Evaluation Planning

All the panelists emphasized the importance of building in evaluation from the start of program planning. It is important to take a careful look at your program and the STEM ecosystem around it to understand and inform the appropriate evaluation (i.e., articulate the logic model for the program). Having this kind of strategy in place helps focus evaluations to implement change in sensible and meaningful ways. For most programs it can be difficult or impossible to get rid of assumptions and mitigating factors, but clearly defining the desired program effect early is helpful. Once this careful consideration has been done,
the evaluator can look at the range of evaluation options available (by doing a literature search or other type of benchmarking) and pick the one that best fits the criteria.

To make sure that an evaluation does not detract from an event, consulting all the stakeholders early is essential. A primary issue, discussed in more detail below, is how the evaluation data will be used. For example, if someone running a program has no plan to use the data, then collecting the data will be a burden without a benefit. However, properly planned evaluation can be used to demonstrate return on investment and justify the program. Several panelists discussed the importance of a robust infrastructure for collecting and storing data, which can often be costly. They suggested collaborating with available IT departments to help implement such a system correctly and efficiently. It is also crucial to understand what new data are actually needed; it may be better to prioritize and take advantage of what already exists.

3. Collecting Demographic Data

The audience asked several questions regarding collecting demographic data about participants. The panelists agreed that this is a challenge, but that this information is important to collect. There are efforts to collect these data more broadly at several agencies.

Audience members pointed out that clarity from leadership on the importance of collecting these data would be helpful. There was also a suggestion to reach out to the local diversity office of your agency for help crafting a message to convince leadership that these data are important.

For collecting somewhat sensitive data, understanding the community around a program is important. The audience expressed concern that putting government logos on surveys can depress the response rate, particularly on demographic questions. The panel (and other audience members) responded with encouragement to using surveys that are culturally relevant to the targeted audience, perhaps by leaving off logos. There was further discussion of the importance of building trust with disconnected communities. Several strategies for building trust that came up, including continuous program improvement based on participant feedback; partnering with universities or other trusted organizations to perform the data collection; increasing transparency, through the use of consent forms, so that participants know exactly how their data are going to be used, and annual publications showing how the data are being used.

Several audience members brought up alternative methods for getting approximate demographic data. In particular, websites like schooldigger.com provide information about individual schools that can be used to figure out approximate percentages of various demographic and underrepresented groups. Panelists emphasized that there needs to be more sharing of best practices.
4. Cost Concerns

Members of the audience expressed concern over the potential cost of running an evaluation, which could take away from the ability of agencies to run complete programs. The panelists and other audience members had several strategies for bringing down costs.

The key takeaway to understanding possible costs is to strategically understand the program goals. In particular, once program goals are clearly defined, you can make informed decisions about using an internal (usually less expensive) or external (usually more expensive) evaluator. Further, if there is a strategic approach to program implementation and evaluation that includes guiding questions and theories of action, evaluations can be more focused, thus reducing costs. The panel expressed concern over having some set percentage of program funding automatically designated for evaluation when the cost should really depend on the type of program.

Another method to drive down costs is to leverage resources that already are available. When developing a database, evaluate what IT resources are available. There may also be other sources of general data and shared tools that could lessen some of the evaluation burden. The goal should be to have an infrastructure in place for data collection and evaluation before evaluation to reduce costs. If the structure is not already in place, performing benchmarking of other evaluation efforts and literature searches to identify evaluation structures and resources can also save money in the long run.

Finally, another method to bring down costs is to evaluate a program using only a sample of participants. Especially for an evaluation early in a program’s life cycle, evaluating limited numbers of participants can provide an affordable way to streamline the evaluation process before rolling it out to the entire program. In some cases, evaluating only a sample can provide the information that is needed.

F. Breakout Groups: Discuss Evaluation Issues

The workshop continued with a session of breakout groups intended to serve two broad purposes: (1) allow people from different backgrounds and roles to hear about different perspectives on STEM program evaluation, and (2) begin building an ongoing community of practice among the STEM community from different agencies with common interests.

To serve these purposes, the members of each breakout group were preselected from registered attendees. The workshop organizers assigned group members to maximize the number of perspectives within a group. Each group consisted of people from different agencies and different roles in STEM evaluation (frontline STEM providers, managers, supervisors, funders, and evaluators) to avoid having people from the same organization or in similar roles together.
To focus the discussion, the breakout groups were provided an hour to discuss four questions:

1. What challenges have you seen in your program evaluations?
2. What solutions have you implemented for those challenges?
3. What are the benefits you see from performing evaluations?
4. What are some of your organizations’ accomplishments in STEM evaluation?

These questions provided open-ended guidance to the breakout groups. The idea was to start a discussion by revealing the challenges that different group members were having with program evaluation and then discuss strategies for solutions. The final two questions allowed the participants to discuss positive outcomes from their evaluations that could motivate communities of interest after the workshop. The questions functioned as a transition from discussions earlier in the morning. Immediately before the breakout group, a panel on evaluations outside DoD addressed challenges to program evaluation and strategies for solutions; many of the breakout groups picked up and continued those lines of discussion in an effort to begin establishing a community of practice.

Each breakout group began by having participants introduce themselves with their name, organization, title, and STEM role relating to evaluation. Due to time constraints and spirited discussions, not all groups were able to address all questions in full; there were fewer answers to the third and fourth questions than to the first two. Most of the groups continued their discussion into lunch.

The following summary of the breakout groups’ discussions represents a consolidation without attribution. Attention was taken to accurately reflect the information discussed and avoid adding any reviewer opinions/commentary. The ultimate goal is to showcase the ideas and viewpoints shared across the STEM community. Within each question, the points are ordered roughly by how common they were. In other words, discussion points that occurred in multiple groups are presented before points that were brought up in only one group.

1. **What challenges have you seen in your program evaluations?**

The main themes from the breakout groups regarding STEM evaluation challenges can be summarized into four main points: (1) finding resources to do evaluations, (2) explaining the need and motivation for data collection and the type of data being collected, (3) creating a standard data-collection approach, and (4) acquiring analyzable data. The discussion surrounding these main points is summarized below, followed by additional challenges that some groups identified but were not common across multiple discussions.

Echoing points raised in the panel, many groups emphasized that evaluation planning requires resources (e.g., time, money, and people), and the challenge is allocating these...
resources from the start as part of STEM program planning to ensure resource availability—once a program begins, resources are set (e.g., limited funds) for evaluation. So, unless a program has planned for evaluation and necessary resources from the start, then collecting the data deemed necessary by management creates a burden for those actually doing the work.

The next three commonly shared challenges across the breakout groups pertain to different stages of evaluation data: collecting data, ensuring data standards, and acquiring analyzable data. The first challenge regarding data collection is reaching agreement up front about what data will be collected, what makes the data meaningful, and in what ways the data will be used. A focus for data collection needs to be on asking the right questions (planning strategy) and putting the data in context. This means collecting only the data that are needed, aiming to achieve strategic relevance with the shortest and simplest data collection possible.

Data standards are an additional challenge. Primarily, the issue is that authorities and responsibilities for establishing data standards are scattered (and vary) across STEM organizations. In addition, it is difficult to develop a single unified data-collection approach that suits everyone’s needs; the STEM evaluation community should strive to balance common data collection with the nuanced needs of each program. Finally, a commonly shared challenge is acquiring appropriate data for analysis: in essence, the challenge is acquiring analysis-ready data that could guide the building of datasets to be used in the future. This would help avoid the high costs caused by lack of data structure (i.e., needing to spend resources creating and acquiring data in a usable format for analysis).

Breakout group members discussed two additional issues related to data collection and issues related to STEM awareness and trends. Regarding data collection, some additional discussions focused on the challenges associated with reporting short-term return on program investment when tracking students is not permitted. Another challenge is managing data calls: data calls and requests from different offices are not well communicated, and changes in direction and scope from management present problems for focused data collection.

Finally, how to provide STEM awareness to communities is a lasting challenge that is also closely tied to STEM trends. Determining trends in STEM over time is a challenge because of population shifts that change demographics, along with immigration issues that may make students and parents reluctant to give data.

2. What solutions have you implemented for those challenges?

Overall, proposed solutions to challenges consisted of strategies that have already been implemented or are currently being planned. Common actionable solutions included

(1) implementing more strategic evaluation planning from the ground up to overcome
challenges (i.e., at the program planning stage), (2) establishing and understanding an appropriate evaluation context for data collection specific to a program, and (3) defining program terms and common data metrics at the program manager and stakeholder level.

Before establishing a data-collection strategy for program evaluation, a program needs to determine and implement an overall evaluation plan. Collecting the appropriate data starts with an evaluation plan that is strategically focused on collecting data aligned with the kind of change the program wants to implement. Building this evaluation plan takes time, but ultimately helps with showing the value of programs.

Following an evaluation plan, establishing a data-collection strategy is an actionable solution aimed at collecting the kind of data a program needs. Evaluation efforts should start small with available funds and a justified need; the scope of evaluation should increase only with additional funds. Establishing a justified need for data collection that’s couched in an appropriate evaluation context can help focus data metrics. That is, consider the entire ecosystem that a program exists in to determine the context for evaluation and to guide data interpretation. For example, the criteria needed to evaluate remote and insolated installations are different from the criteria needed to evaluate larger programs.

A related solution to establishing a data collection plan focuses on defining program terms and data structure before data collection. Implementing an evaluation and data-collection plan starts with stakeholders and program managers agreeing on program definitions and establishing common data structures and metrics focused on the program’s overall evaluation plan.

Additional solutions discussed by some groups included tailoring evaluations, communicating successes, and reducing cost. With regard to tailoring evaluations, one must understand the landscape that a program functions in and tailor evaluations to the program’s specific community and cultural needs instead of forcing the same evaluation approach on all programs. In terms of evaluation communication, think of three “R’s”—Relevance, Response, and Results—and try to fit descriptions into 200 words with the use of pictures for the purposes of PR or testimonies of program success. Finally, driving down evaluation costs could be done by taking advantage of data already available from organizations such as the National Academies, International Education Society (IES), and National Center for Education Statistics (NCES) if they are in a usable format. In other words, draw upon the existing research base for tools and data; do not try and create boutique resources for each program.

3. **What are the benefits you see from performing evaluations?**

   Overall, common benefits from performing program evaluations include: (1) iteratively improving programs, (2) maximizing outcomes, and (3) providing strategic support.
Evaluations help to improve programs as part of an evolutionary (iterative) process by helping evaluators and managers understand the baseline of where a program currently stands, determine how to modify programs, assess the results of changes, and decide what new programs to pilot. Evaluations aid in understanding practices that work and maximizing these practices by implementing them in other programs, potentially with limited resources. In addition, collecting appropriate data helps with understanding what data are needed and how much is necessary.

As part of the iterative process, evaluation also helps demonstrate how a program is meeting an organization’s strategic plan or mission and helps align STEM programs with higher-level strategic goals; overall this helps to defend and justify programs by providing strategic support. A side benefit of this process is that individuals and stakeholders will reinvest in the program when they see it is trying to improve.

4. What are some of your organizations’ accomplishments in STEM evaluation?

Overall, accomplishments from STEM evaluation include (1) sharing of evaluation tools and findings, and (2) transparency by documenting accomplishments and progression.

Sharing the outcome of a program evaluation with the larger STEM community allowed programs to draw information about what tools, lessons, development, and design are in the STEM community and successful now. In a similar light, evaluation enabled one program to document accomplishments and show the progression of students from one program to another, ultimately highlighting the STEM portfolio’s ability to maintain interest of students over time.

Finally, two other accomplishments were (1) evaluations acting as a program filter and (2) evaluations helping show the development of social skills. As a program filter, one program noted that evaluation data were used to judge geographic diversity and to effectively filter out programs that were not meeting expectations. Regarding social development, one particular accomplishment example was a program showing that a survey of attitudes helped evaluate the development of social skills.

G. Program Evaluation Example: STARBASE

Toward the workshop’s goal of exposing attendees to high-quality DoD STEM evaluations, Gary Behrens of FifthTheory and Cyndy Mickevicius from The Spectrum Group discussed the ongoing evaluations that they conduct on the STARBASE program. Mr. Behrens’ talk began with an overview of the STARBASE program, was followed by a detailed discussion of the evaluations performed, and finished with some lessons learned.
1. **An Overview of STARBASE**

   The STARBASE program, run by the Office of the Assistant Secretary of Defense Manpower and Reserve Affairs (OASD/M&RA), has been ongoing for 25 years. The program, which targets Title I schools, provides fifth-grade students with 25 hours of STEM instruction and activities. These students go to 1 of 66 military-hosted STARBASE academies at locations across the U.S. and in Puerto Rico, where instructional staff from the academy give STEM concept exposure using an inquiry-based curriculum.

2. **The Evaluation**

   The main goal of the STARBASE evaluation is to test how well the program is meeting its objectives. The evaluators collaborated with stakeholders to determine what was important and to design the assessment. The evaluation team worked to control alternative explanations for effects but could not control everything by performing true random control trials. Instead, the evaluation team conducts a mixed-method assessment on one STARBASE class at each location per year and aggregate data analyses.

   First are pre- and post-program measures of student knowledge and attitudes. The knowledge test is designed to measure a general understanding of STEM concepts (matching STARBASE curriculum objectives). The student attitude survey focuses on STEM awareness, interest and vocational aspirations, science learning confidence, military impressions, and STARBASE. The content of both tools is kept relatively stable over years (modified by ~10%) to allow for cross-year comparisons. Quantitative analysis is mostly done on the shifts from pre-program scores to post-program scores. Results show that student performance on the STEM knowledge test consistently improves 25%–30%, and on the STEM attitudes survey over 80% of items increase significantly in favorability ratings.

   Another component of the evaluation is a post-program survey of teacher opinions. This survey focuses on perceptions of STARBASE impact on students, teachers, and the classroom environment. The quantitative results of the student surveys as well as the qualitative results of the teacher survey are described in a comprehensive final report to DoD and location-specific summaries are also reported to the academies on an annual basis.

---

4 Title I schools are schools that receive federal assistance because they serve high percentages of children from low-income families. See https://www2.ed.gov/programs/titleiparta/index.html.

5 Randomized controlled trials would require that some students not receive the instruction (i.e., control group) so that they could be compared to those who did.
3. Lessons Learned

The following key lessons resulted from the STARBASE evaluation:

- The DoD STARBASE program is achieving its objective of fostering greater student interest in STEM learning and occupations.
- It is impossible to over-plan—establish a project plan, a realistic schedule, and a plan for thorough quality assurance check of evaluation results.
- Communicate often—check in with project team, program staff, and vendors often; respond to field user concerns quickly.
- Keep innovating—while it is important to be consistent year to year, it is also key to keep up with program improvements and bring fresh insights to each annual report.

H. Portfolio Evaluation Example: Army Education Outreach Program

1. Overview

In an effort to showcase evaluation capabilities, the next presentation addressed the portfolio level of analysis of STEM efforts, that is, evaluating multiple programs that are part of a combined initiative. AEOP is a portfolio of multiple STEM programs, including STEM competitions and programs for students from elementary to high school, apprenticeships for high school and college students, and STEM-focused professional development for teachers. Dr. Carla Johnson and Dr. Toni Sondergeld presented the scope of their ongoing evaluation efforts for this portfolio, including both short- and long-term studies looking at student and teacher outcomes. A total of 11 programs were evaluated for FY18, with survey responses from 3,613 students and 707 adults. Data also came from interviews/focus groups from 300+ participants and 12 site visits.

2. Evaluation

The suite of evaluations was aligned with three AEOP priorities: (1) broaden, deepen, and diversify the pool of STEM talent in support of DoD; (2) support and empower educators with unique Army research and technology resources; (3) develop and implement a cohesive, coordinated, and sustainable STEM education outreach infrastructure across the Army. With these priorities in mind, a centralized evaluation plan guided research questions regarding STEM learning/competency gains, repeated participation among program alumni, and attitudes toward, and awareness of, STEM and Army/DoD STEM. A combination of short- and long-term questions allowed understanding how programs are performing within a given year and longitudinally. Dr. Johnson also mentioned the new Federal STEM strategic plan as a point of reference when planning evaluations.
The research questions are tied directly to outcomes, which are tied directly to measures. These measures consist of 24-item questionnaires for the student participants, questionnaires for the program mentors, focus groups that include both of these populations, surveys and interviews with program alumni, interviews with program directors, and site visits. Program materials are also reviewed, and in the interest of promoting evidence-based practices, best practices are disseminated across the range of portfolio programs.

All AEOP programs adhere to a master evaluation calendar and use the Cvent online registration system, which allows collection of detailed participant data. A participant is seen as underserved by self-identified membership in two or more of these categories: students with disabilities; students with English as a second language; first-generation college students; students in rural, frontier, or other federally targeted outreach schools; or females in certain STEM fields.

The AEOP suite of evaluations are notable in their recommendations, given that the annual reports for each program tie findings directly to actionable recommendations, and subsequent reports from following years describe actions taken to address previous recommendations and the progress then seen. Dr. Johnson called this “moving beyond lip service.” In addition to these individual program reports, a summative portfolio evaluation report is compiled annually, with similarly clear recommendations.

Dr. Johnson also stressed challenges facing AEOP evaluation, including timing of data collection, curation and analysis of participation data, participation rates for evaluation surveys, and logistics for site visits. An audience member brought up a different challenge, citing a statistic describing the percentage of minority STEM majors who either switch majors or drop out (30%) and asking what should be done about that number. Dr. Johnson pointed out a potential reason for this STEM attrition, saying that these students often do not have the support they need and that it is a matter of getting the resources and support to prevent the STEM pipe leakage.

3. Lessons Learned

The following are the key takeaways from AEOP evaluation:

- Utilize a combination of short- and long-term research questions to understand progress.
- Use a centralized registration and reporting system to ease burden across programs.
- Leverage evaluation findings as a mechanism for continuous improvement by providing recommendations for program development; it is not enough to just state findings.
• Determine program responsiveness to previous evaluation findings.

I. Principle Evaluation Example: Pipeline Persistence

1. Overview

The third level of analysis of DoD STEM Evaluation and Assessment Capability is a principle evaluation. This presentation focused on the principle of keeping students engaged in STEM as they progress through schooling. Dr. Adam Maltese, Indiana University, in his talk “Example of a Principle Evaluation: Pipeline Persistence” emphasized three undergirding principles for evaluating STEM persistence characterized here as follows:

1. Engage frequently in iterative cycles of quantitative and qualitative evaluations.
2. Get multiple perspectives (e.g., young and old; STEM and non-STEM) to avoid biases.
3. Realize that how STEM participants view what happened usually is more important than reality.

Dr. Maltese’s presentation illustrated these principles using examples and data beginning with an observation about multiple perspectives. He said that youth participants provide real-time information about STEM participation but may lack perspective on longer term impact of events; adults can provide perspective tempered by even false memories of previous events.

2. Evaluation

Dr. Maltese’s repeated cycles of quantitative and qualitative evaluations show that interest in STEM is due to a wide variety of events and influences by different people, including the individuals themselves, their parents, and teachers. Different racial and ethnic groups gain interest in STEM at different times in school, but it is unclear why. One of the few clear findings is that male and female career choices reflect traditional gender roles. That is, females are more interested in medicine—in helping animals and people—with consistently less interest, regardless of grade level, in science and engineering; males are the reverse. Differences also are evident in what principal type of experience sparked STEM interest. Females pointed equally to good grades in STEM courses and to math problems, logic games, pattern tests. In contrast, males cited building, tinkering, taking apart mechanical objects or electronics and, to a lesser extent, books, magazines, or comics.

A research focus at Indiana University has been on the role of making things and of the opportunity for failure on fostering a STEM pathway. The premise is that failure is an important part of science because it teaches how to overcome problems. Yet education does
not offer that opportunity except perhaps in graduate school. Failure is part of the process of making things better. Adults are afraid of failure for their children; children are more willing to accept failure. Parents can reduce their fear by fostering problem-solving through design and making things. It mostly helps with the engineering part of STEM. Unfortunately, designing and building is a White middle class activity, which children from underserved communities do not often get to do.

Maltese’s closing thought was to reemphasize the importance of using both qualitative and quantitative approaches to STEM evaluation. We need interviews and observations to supplement the quantitative data we collect. For example, ask fifth graders to define engineering, technology, or science. The child offers thoughts such as math, innovation, and help make the world a better place. Maltese asked, does it matter how adults, who influence STEM interests, define them? At the same time, we need to collect data with surveys, from academic performance—traditional grades and performance-based assessments—and career interest profiles, among others. We need to combine frequent data collection, multiple perspectives, and participant views to understand the STEM pipeline.

3. Lessons Learned

The following are the key takeaways from the pipeline persistence evaluation:

- Persistence of a student in the pipeline is influence by a range of people: the student themselves, peers, parents, and teachers.
- Some differences in persistence may be correlated with demographics.
- Understanding persistence influences and the timing of persistence decisions may help in increasing persistence.

J. STEM Workforce Program Evaluation Example: SMART Scholarship

1. Overview

In keeping with Federal and DoD STEM Strategic Plan’s emphasis on STEM pipeline and workforce development, Dr. Asha Balakrishnan from the Science and Technology Policy Institute (STPI) of the Institute for Defense Analyses (IDA) presented on the program evaluation conducted on the Science, Mathematics, and Research for Transformation (SMART) scholarship for service program. SMART is a DoD program that provides scholarships for students pursuing a bachelor’s, master’s, or doctoral degree in one of 19 STEM disciplines. From 2006 to 2016, the majority of the 1,962 awardees received engineering and computer science degrees, with 45% obtaining a BS, 25% an MS, and 30% a PhD. As legislatively mandated by 10 U.S. Code § 2192a, “The Secretary of Defense shall carry out a program to provide financial assistance for education in science,
mathematics, engineering, and technology skills and disciplines that, as determined by the Secretary, are critical to the national security functions of the Department of Defense and are needed in the Department of Defense workforce.” To that end, recipients are awarded full tuition and a stipend, along with a summer internship position at their sponsoring DoD facility. For each year of scholarship, scholars commit to working 1 year as civilian in DoD laboratories, research and development centers, and facilities. There are two types of scholarships. One is “recruitment,” for people who are not currently employed by the government and are attending college (undergraduate or graduate school). The other is “retention,” for those currently employed by DoD labs or facilities who would like to pursue additional schooling. Overall, 87% of the awards were recruitment scholars, and 13% were retention scholars. As a point of reference, DoD science and engineering (S&E) workforce in 2016 numbered 140,000, indicating that SMART scholars represent a relatively small percentage of the overall S&E workforce.

2. Evaluation

The SMART program within OUSD(R&E) Laboratories and Personnel Office tasked IDA to conduct an in-depth evaluation that included a “process evaluation” by analyzing the SMART Program’s past and present program-management processes and an “outcome/impact evaluation” of the SMART program. In addition, the evaluation was designed to lead to recommendations for an ongoing evaluation plan for the SMART program for continuous improvement and tracking progress.

The overall evaluation included multiple sources of data and analytic methods. Primarily for the process evaluation program documentation, legislation, and administrative records were reviewed. For both the process and outcome evaluation components, interviews were conducted with over 230 stakeholders, including those at the SMART program office, S&T managers and human resource personnel at labs/facilities that were hiring the SMART scholars, scholars at all levels of progression through the program and at multiple facilities, Service liaisons, contract support for the SMART program, and academic advisors of the scholars in PhD programs. In addition, a survey that was developed and administered to all past and present scholars (administered in Winter 2017 to scholars in 2005 to 2015 cohorts) led to 1,112 survey responses (65% response rate). Finally, the personnel records of the scholars held by the Defense Manpower Data Center were analyzed, along with a matched comparison group of S&T workers (i.e., similar positions, similar facilities, and similar hiring timeframe).

Findings from the process evaluation include how the program has steadily increased in size over time, except for a few years when funding was reduced from the prior year. The increase in program size is an indication of how the participants perceive the program. Reduction of funding from one year to the next is difficult for the program to adapt since most scholars are receiving scholarships for multiple years. A relatively modest reduction
in funding leads to a significant drop in the number of new scholarships that can be awarded because of the existing scholars who need to be funded first before new scholars can be added to the program. One process observed was that scholars were assigned to a specific facility (i.e., their ultimate place of employment if they continue in the program) at the time of scholarship award (i.e., before the scholar or the facility really have understood one another). This has benefits and challenges. A sustained relationship between the scholar and the facility is developed over time so that after graduation, when the scholar is hired, they can contribute immediately. However, the DoD stakeholders may experience difficulties aligning the “fit” of the scholar with a specific facility so early in the process and predicting workforce needs years into the future, given that the average length of a scholarship is 2 years with payback employment lasting for an additional 2 years after graduation.

Findings from the outcome evaluation indicate that the SMART scholars were on average of higher quality than the comparison group of S&E personnel hired by DoD through mechanisms other than the SMART program. Indicators of higher quality were based on the interviews with S&E managers, assessment of university quality of the SMART scholars, and promotion rates of recruitment scholars after they were employed (promotion rates of retention scholars were similar to a comparison group).

An analysis of salary indicated that scholars were hired at lower starting salaries, but received raises at a higher rate than comparison groups. This lower starting salary may be problematic in retaining scholars past their commitment, but is similar to other findings of economic principles where people who are required to join an organization (or stay within an organization) may be compensated less than those with a choice.

The retention level of recruitment scholars was less than their comparison group, with about 25% more scholars leaving the facility after 6 years (i.e., 75% retention rate for the comparison group and 50% retention rate for recruitment scholars).

Overall, scholar satisfaction ratings of the program were very high. However, those with lower levels of satisfaction were more likely to have received more years of scholarship (more years of required employment), were more likely to be Ph.D. scholars, were more likely to be at facilities that were relatively farther away from home (before joining the program), and were less likely to state that they enjoyed serving the DoD mission or felt as though they were adequately mentored. Although diversity is not an explicit program goal, analysis of diversity indicated that the SMART scholars were less racially diverse but more gender diverse than the DoD S&E workforce. In addition, a considerable percentage of scholars had not considered DoD employment prior to applying for the scholarship, suggesting that SMART helped attract new talent to the DoD.

The independent third-party evaluation of the SMART program was informative to those managing the SMART program, confirming some of their assumptions as to what
was happening and providing some additional information that they could use to improve the program or as a baseline for future comparison. Some of the recommendations that relate to ongoing efforts to improve the program include addressing diversity by continuing to recruit female scholars while expanding efforts to increase representation of underrepresented minorities, developing activities that may improve retention of high-quality scholars, and continuing to recruit high quality scholars and improve the applicant pool. In addition, some new actions were recommended based on the findings of the evaluation: investigate differences in starting salaries of scholars and work with DoD facilities to understand the salary disparities, and conduct a workforce demand analysis for components and facilities to determine and prioritize the sponsoring facility/laboratory need for SMART scholars, including degree field, degree level, and skill level.

3. Lessons Learned

The following are the key takeaways from the SMART evaluation:

- A variety of data sources and analytic methods provide a depth and breadth of evaluation.
- Leverage evaluation findings to provide targeted recommendations for program improvement (e.g., diversity, recruitment, retention).
- Evaluation findings help assess if a program is meeting goals and functioning as intended.

K. Closing and Next Steps

The workshop had an open discussion at the end of the day to highlight important ideas that were presented throughout the event. Many of the discussed points reiterated key themes from earlier sessions in the workshop. One of the key points was the importance of planning to conduct evaluations. This included the standard planning steps of starting with the goal of the STEM activity to be evaluated and developing an efficient means for collecting and analyzing the appropriate data needed for the evaluation.

The conversation on planning also brought up the point that adequate lead time is required between when new plans are developed and when the plans may actually be implemented. Since many programs are cyclical (annual, semi-annual, or held only during a certain part of the year like summer camps), a change in the evaluation plan may have to be integrated into a STEM program over time and not be a change that can be immediately inserted. If a program is midway through its cycle, it may take until the beginning of the next cycle before a new data collection effort could be implemented.

Another aspect of evaluation planning highlighted was to make sure that the appropriate type and amount of resources were allocated before starting the program because once the program is up and running, it may be difficult to find additional resources
unless they are already part of the plan. Also related to planning and resources was the goal of identifying ways that ongoing or future evaluations can be made more efficient by using resources effectively.

Several attendees expressed appreciation for the discussion throughout the day and reinforced that further development of the evaluation capability within the DoD will be an ongoing process. As a next step, attendees agreed on sharing contact information to promote interaction between members of the community. In addition, attendees requested that the slides of the presentations be shared with the goal of disseminating best practices and materials.
3. Summary

The DoD STEM Evaluation Workshop brought together professionals interested in STEM education programs from multiple DoD agencies (Army, Navy, Air Force, Defense Threat Reduction Agency, Missile Defense Agency, National Security Agency, OUSD Research and Engineering, and OUSD Personnel and Readiness), Federal agencies (Smithsonian, NOAA, NASA, and NSF), academic institutions (U.S. Naval Academy, Indiana University, North Carolina State University, and Drexel University), and a few organizations that provide support to DoD STEM efforts (Great Minds in STEM, American Institute of Research, Spectrum Group, and IDA) to discuss the use of evaluations to strengthen STEM programs and portfolios.

The workshop sessions were organized so that the day started with high-level presentations with broad perspectives by Dr. Lei (Deputy Director of Research, Technology, and Laboratories), who presented the OSD perspective on STEM programs and the importance of evaluation, along with Dr. Werner-Allen (OSTP), who provided an overview of the Federal STEM Strategic Plans along with the Federal perspective. Next, Mr. Lopez provided the OSD STEM Development Office’s perspective on how the DoD STEM community can work together and the value of evaluations for STEM programs. Dr. Diann McCants then presented on the topic of diversity and inclusion to highlight how broadening participation leads to positive outcomes. Then, a panel of STEM professionals from government agencies other than DoD (Dr. Sarah-Kay McDonald from NSF, Dr. John Baek from NOAA, Mr. Rick Gilmore from NASA, and Dr. Carol O’Donnell from the Smithsonian) provided insight into how other government agencies evaluate their STEM programs.

Following the broad perspectives discussed in the morning, a breakout group session was conducted that consisted of dividing the attendees into six groups, where people in a group came from different agencies or had different types of functions in providing or evaluating STEM programs. The group discussions were lively, as indicated by people continuing to discuss their topics during lunch. Some of the discussion points that were similar across groups included the need for planning evaluation as part of planning STEM programs so that the evaluation is aligned with program goals; allocating resources to evaluation; sharing evaluation methods and tools and leveraging other evaluation capabilities; and communication between STEM providers who may be part of the data collection effort, program managers who may shape programs based on evaluations, and organization or agency leaders who may request or use evaluation results to make decisions.
After lunch, four presentations were given to address a range of evaluation types (i.e., evaluation of a single program for children, evaluation of a program designed for workforce development, evaluation of a portfolio of programs, and evaluation of a concept or principle that can inform program development. Gary Behrens and Cyndy Mickevicius provided an overview of the STARBASE program and described their annual evaluation process, which includes a knowledge assessment to demonstrate how children are learning STEM content from their STARBASE experience. Dr. Carla Johnson and Dr. Toni Sondergeld presented about their ongoing evaluation of the set of programs that are part of the AEOP portfolio. The program evaluations assess how the AEOP mission is accomplished across the portfolio of programs. Dr. Adam Maltese described the evaluations that informed the principle of “pipeline persistence,” which is focused on keeping students engaged in STEM education as they progress through school. Dr. Asha Balakrishnan described the method used and presented the findings of an in-depth evaluation of the SMART scholarship for service program, which were that the program attracts high-quality talent to DoD but there are some issues with retention of that talent. This SMART evaluation has led to some program changes as part of a continuous-improvement effort. The workshop concluded with a period of open discussion, where attendees reiterated many of the main points from the workshop.

After the workshop, follow-up efforts were made to continue to foster a community of practice. As part of that process, several steps were taken immediately after the workshop. All workshop participants received a list of attendees and their e-mail addresses (see Appendix B) to facilitate the creating communities of practice. In addition, the slides from the presentations used by the workshop’s speakers were shared with the attendees. Several attendees followed up to convey satisfaction with the effectiveness of the workshop and pass along resources relevant to the discussions.

Further work is underway to continue building on the goals provided at the beginning of this report. There are plans to create opt-in mailing lists around various topics of interest to allow attendees to connect within more targeted communities. The slides from the workshop as well as a few other resources will form the starting point of an evaluation best practices repository. Eventually, this repository will be built out with a variety of sources that evaluators and planners can use to follow through on some of the ideas presented in the workshop. Finally, this report on the evaluation workshop will be distributed to the attendees and other interested parties so that the lessons of the workshop are archived and distributed.

The workshop can benefit DoD STEM in another way by providing a set of examples that can be used in the development of a repository of best practices for STEM evaluation. The four examples of evaluations (STARBASE, AEOP, SMART, and pipeline of persistence) cover the three levels of effort in DoD STEM’s evaluation and assessment
capability plans. The repository can grow as more high-quality evaluations are conducted, and it will become a resource to benefit future STEM evaluation.
Appendix A.
Planned Workshop Agenda

The table below shows the planned agenda for the workshop. During the workshop, the start times were adjusted as needed; due to time constraints, the second breakout group session was replaced by an open discussion session.

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity/Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00</td>
<td>Registration; coffee and pastries</td>
</tr>
<tr>
<td>8:25</td>
<td>Welcome and Logistics</td>
</tr>
</tbody>
</table>
| 8:30   | Keynote Address: The Importance of Evaluations for STEM Success  
Dr. JihFen Lei, Dep. Dir. for Research, Technology and Laboratories, OUSD (R&E) |
| 8:55   | Federal STEM Context and FC STEM Strategic Plan  
Dr. Jon Werner-Allen, Office of Science and Technology Policy |
| 9:10   | OSD STEM Perspective  
Mr. Louie Lopez, Director, STEM, OUSD(R&E) |
| 9:25   | Diversity and Inclusion Considerations in Evaluations  
Dr. Diann McCants, Contractor Support, OUSD(R&E) |
| 9:40   | Networking break; coffee and refreshments |
| 9:55   | Panel: Discussion of Evaluations outside of DoD  
Dr. Sarah-Kay McDonald (NSF), Mr. Richard Gilmore (NASA), Dr. Carol O'Donnell (SI),  
Dr. John Baek (NOAA) |
| 11:00  | Breakout Groups: Discuss Evaluation Issues  
Groups respond to list of questions about STEM evaluations |
| 12:00  | Networking Lunch |
| 13:00 (30 minutes) | Example of Program Evaluation: STARBASE  
Ms. Cyndy Mickevicius (Spectrum Group) and Mr. Gary Behrens (Fifth Theory) |
| 13:30  | Example of Portfolio Evaluation: AEOP  
Dr. Carla Johnson, North Carolina State University and  
Dr. Toni Sondergeld, Drexel University |
| 14:00  | Networking break; coffee and refreshments |
| 14:15  | Example of Principle Evaluation: Pipeline Persistence  
Dr. Adam Maltese, Indiana University |
| 14:45  | Example of Workforce-related Evaluation: SMART Scholarship  
Dr. Asha Balakrishnan, IDA |
| 15:15  | Breakout Groups: Topics of Interest  
Groups organized by topics of interest to focus on building communities of interest |
| 16:15  | Summary/Concluding Remarks  
IDA |
Appendix B.
List of Participants

<table>
<thead>
<tr>
<th>Last Name</th>
<th>First Name</th>
<th>Email address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baek</td>
<td>John</td>
<td><a href="mailto:john.baek@noaa.gov">john.baek@noaa.gov</a></td>
</tr>
<tr>
<td>Balakrishnan</td>
<td>Asha</td>
<td><a href="mailto:abalakri@ida.org">abalakri@ida.org</a></td>
</tr>
<tr>
<td>Behrens</td>
<td>Gary</td>
<td><a href="mailto:gary.behrens@FifthTheory.com">gary.behrens@FifthTheory.com</a></td>
</tr>
<tr>
<td>Belanich</td>
<td>Jim</td>
<td><a href="mailto:jbelanich@ida.org">jbelanich@ida.org</a></td>
</tr>
<tr>
<td>Bierman</td>
<td>Mervin</td>
<td><a href="mailto:mwBierm@nsa.gov">mwBierm@nsa.gov</a></td>
</tr>
<tr>
<td>Bowens</td>
<td>Victoria</td>
<td><a href="mailto:victoria.l.bowens.civ@mail.mil">victoria.l.bowens.civ@mail.mil</a></td>
</tr>
<tr>
<td>Britt-Crane</td>
<td>Michael</td>
<td><a href="mailto:michael.d.britt-crane.civ@mail.mil">michael.d.britt-crane.civ@mail.mil</a></td>
</tr>
<tr>
<td>Brown</td>
<td>Carl</td>
<td><a href="mailto:carl.r.brown2.civ@mail.mil">carl.r.brown2.civ@mail.mil</a></td>
</tr>
<tr>
<td>Bullock</td>
<td>Amanda</td>
<td><a href="mailto:amanda.bullock.3@us.af.mil">amanda.bullock.3@us.af.mil</a></td>
</tr>
<tr>
<td>Cantu</td>
<td>Andrea</td>
<td><a href="mailto:andrea.v.cantu.ctr@mail.mil">andrea.v.cantu.ctr@mail.mil</a></td>
</tr>
<tr>
<td>De La Cruz</td>
<td>Denise</td>
<td><a href="mailto:denise.b.delacruz.civ@mail.mil">denise.b.delacruz.civ@mail.mil</a></td>
</tr>
<tr>
<td>Desjardins</td>
<td>Candi</td>
<td><a href="mailto:candida.desjardins@navy.mil">candida.desjardins@navy.mil</a></td>
</tr>
<tr>
<td>Dobbins</td>
<td>Christian</td>
<td><a href="mailto:cdobbins@ida.org">cdobbins@ida.org</a></td>
</tr>
<tr>
<td>Druggan</td>
<td>Tom</td>
<td><a href="mailto:Tom.Druggan@mda.mil">Tom.Druggan@mda.mil</a></td>
</tr>
<tr>
<td>Evans</td>
<td>LaShauna</td>
<td><a href="mailto:lashauna.c.evans.ctr@mail.mil">lashauna.c.evans.ctr@mail.mil</a></td>
</tr>
<tr>
<td>Fedele</td>
<td>Emily</td>
<td><a href="mailto:efedele@ida.org">efedele@ida.org</a></td>
</tr>
<tr>
<td>Feichtl</td>
<td>Kenneth</td>
<td><a href="mailto:Kenneth.R.Feichtl@nga.mil">Kenneth.R.Feichtl@nga.mil</a></td>
</tr>
<tr>
<td>Gilmore</td>
<td>Richard</td>
<td><a href="mailto:richard.l.gilmore@nasa.gov">richard.l.gilmore@nasa.gov</a></td>
</tr>
<tr>
<td>Hall</td>
<td>Katie</td>
<td><a href="mailto:catherine.m.hall.civ@mail.mil">catherine.m.hall.civ@mail.mil</a></td>
</tr>
<tr>
<td>Heckman</td>
<td>David</td>
<td><a href="mailto:david.n.heckman.ctr@mail.mil">david.n.heckman.ctr@mail.mil</a></td>
</tr>
<tr>
<td>Heilman</td>
<td>JC</td>
<td><a href="mailto:joyce.c.heilman.civ@mail.mil">joyce.c.heilman.civ@mail.mil</a></td>
</tr>
<tr>
<td>Johnson</td>
<td>Carla</td>
<td><a href="mailto:cjohns26@ncsu.edu">cjohns26@ncsu.edu</a></td>
</tr>
<tr>
<td>Johnson</td>
<td>Clarence</td>
<td><a href="mailto:clarence.a.johnson20.civ@mail.mil">clarence.a.johnson20.civ@mail.mil</a></td>
</tr>
<tr>
<td>Kent</td>
<td>Evelyn</td>
<td><a href="mailto:evelyn.w.kent.civ@mail.mil">evelyn.w.kent.civ@mail.mil</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Last Name</strong></th>
<th><strong>First Name</strong></th>
<th><strong>Email address</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolodrubetz</td>
<td>Dan</td>
<td><a href="mailto:dkolodru@ida.org">dkolodru@ida.org</a></td>
</tr>
<tr>
<td>Laurendine</td>
<td>Lisa</td>
<td><a href="mailto:lisa.laurendine@mda.mil">lisa.laurendine@mda.mil</a></td>
</tr>
<tr>
<td>Lopez</td>
<td>Louie</td>
<td><a href="mailto:louie.r.lopez.civ@mail.mil">louie.r.lopez.civ@mail.mil</a></td>
</tr>
<tr>
<td>Maltese</td>
<td>Adam</td>
<td><a href="mailto:amaltese@indiana.edu">amaltese@indiana.edu</a></td>
</tr>
<tr>
<td>McCants</td>
<td>Diann</td>
<td><a href="mailto:diann.w.mccants.ctr@mail.mil">diann.w.mccants.ctr@mail.mil</a></td>
</tr>
<tr>
<td>McDonald</td>
<td>Sarah-Kay</td>
<td><a href="mailto:sarmcdon@nsf.gov">sarmcdon@nsf.gov</a></td>
</tr>
<tr>
<td>McKittrick</td>
<td>Alexis</td>
<td><a href="mailto:amckittr@ida.org">amckittr@ida.org</a></td>
</tr>
<tr>
<td>Mellado</td>
<td>Ray</td>
<td><a href="mailto:rmellado@greatmindsinstem.org">rmellado@greatmindsinstem.org</a></td>
</tr>
<tr>
<td>Michalopoulos</td>
<td>Christos</td>
<td><a href="mailto:christos.michalopoulos@noaa.gov">christos.michalopoulos@noaa.gov</a></td>
</tr>
<tr>
<td>Mickevicius</td>
<td>Cyndy</td>
<td><a href="mailto:cmickevicius@spectrumgrp.com">cmickevicius@spectrumgrp.com</a></td>
</tr>
<tr>
<td>Miranda</td>
<td>Kathleen</td>
<td><a href="mailto:kathleen.miranda@navy.mil">kathleen.miranda@navy.mil</a></td>
</tr>
<tr>
<td>Molina</td>
<td>Jess</td>
<td><a href="mailto:jessica.a.molina2.ctr@mail.mil">jessica.a.molina2.ctr@mail.mil</a></td>
</tr>
<tr>
<td>Moran</td>
<td>Angela</td>
<td><a href="mailto:amoran@usna.edu">amoran@usna.edu</a></td>
</tr>
<tr>
<td>Moses</td>
<td>Frank</td>
<td><a href="mailto:fmoses@ida.org">fmoses@ida.org</a></td>
</tr>
<tr>
<td>O’Donnell</td>
<td>Carol</td>
<td><a href="mailto:ODonnellC@si.edu">ODonnellC@si.edu</a></td>
</tr>
<tr>
<td>O’Toole</td>
<td>Michael James</td>
<td><a href="mailto:michael.j.otoole.civ@mail.mil">michael.j.otoole.civ@mail.mil</a></td>
</tr>
<tr>
<td>Pamulapati</td>
<td>Jagadeesh</td>
<td><a href="mailto:jagadeesh.pamulapati.civ@mail.mil">jagadeesh.pamulapati.civ@mail.mil</a></td>
</tr>
<tr>
<td>Park</td>
<td>Anna</td>
<td><a href="mailto:apark@greatmindsinstem.org">apark@greatmindsinstem.org</a></td>
</tr>
<tr>
<td>Pena</td>
<td>Vanessa</td>
<td><a href="mailto:vpena@ida.org">vpena@ida.org</a></td>
</tr>
<tr>
<td>Poland</td>
<td>Susan</td>
<td><a href="mailto:susan.m.poland@nasa.gov">susan.m.poland@nasa.gov</a></td>
</tr>
<tr>
<td>Simmons</td>
<td>Andrea</td>
<td><a href="mailto:andrea.e.simmons.ctr@mail.mil">andrea.e.simmons.ctr@mail.mil</a></td>
</tr>
<tr>
<td>Simpson</td>
<td>Michael</td>
<td><a href="mailto:Michael.M.Simpson1@Navy.mil">Michael.M.Simpson1@Navy.mil</a></td>
</tr>
<tr>
<td>Sondergeld</td>
<td>Toni</td>
<td><a href="mailto:tonis519@aol.com">tonis519@aol.com</a></td>
</tr>
<tr>
<td>Steinbacher</td>
<td>Jeremy</td>
<td><a href="mailto:jeremy.l.steinbacher.ctr@mail.mil">jeremy.l.steinbacher.ctr@mail.mil</a></td>
</tr>
<tr>
<td>Thompson</td>
<td>Kelly</td>
<td><a href="mailto:kelly.c.thompson@navy.mil">kelly.c.thompson@navy.mil</a></td>
</tr>
<tr>
<td>Werner-Allen</td>
<td>Jon</td>
<td><a href="mailto:Jonathan.W.Werner-Allen@ostp.eop.gov">Jonathan.W.Werner-Allen@ostp.eop.gov</a></td>
</tr>
<tr>
<td>Williams</td>
<td>Reginald</td>
<td><a href="mailto:reginald.g.williams@navy.mil">reginald.g.williams@navy.mil</a></td>
</tr>
<tr>
<td>Wilson</td>
<td>Joseph</td>
<td><a href="mailto:jwilson@air.org">jwilson@air.org</a></td>
</tr>
<tr>
<td>Yourick</td>
<td>Debra</td>
<td><a href="mailto:debra.l.yourick.civ@mail.mil">debra.l.yourick.civ@mail.mil</a></td>
</tr>
</tbody>
</table>
### Appendix C.
**Diversity and Inclusion in STEM Evaluation: Quick Knowledge Check**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.</strong></td>
<td>Research has shown that diversity among top leaders and problem-solvers is critical to fostering creativity and innovation in STEM.</td>
<td>T</td>
</tr>
<tr>
<td><strong>2.</strong></td>
<td>Diversity has little impact on a company’s financial profitability.</td>
<td>T</td>
</tr>
<tr>
<td><strong>3.</strong></td>
<td>Women are more likely to leave the STEM workforce for family concerns as opposed to workplace conditions.</td>
<td>T</td>
</tr>
<tr>
<td><strong>4.</strong></td>
<td>Even with a systematic approach to program evaluation, all stages of evaluation design and execution are open to bias.</td>
<td>T</td>
</tr>
<tr>
<td><strong>5.</strong></td>
<td>Measuring concepts such as a STEM identity and sense of belonging can provide insight into whether a student will persist on a STEM pathway.</td>
<td>T</td>
</tr>
<tr>
<td><strong>6.</strong></td>
<td>Looking at 50 years of research on the Draw-a-Scientist Test, children’s depictions of scientists have become more gender diverse over time, but children still associate science with men as they grow older.</td>
<td>T</td>
</tr>
<tr>
<td><strong>7.</strong></td>
<td>Disaggregating data by factors such as STEM discipline, gender, race and ethnicity is not especially useful because you are dealing with such small sample sizes.</td>
<td>T</td>
</tr>
</tbody>
</table>
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>A&amp;S</td>
<td>Acquisition and Sustainment</td>
</tr>
<tr>
<td>AAAS</td>
<td>Association for the Advancement of Science</td>
</tr>
<tr>
<td>AEOP</td>
<td>Army Education Outreach Program</td>
</tr>
<tr>
<td>AT&amp;L</td>
<td>Acquisition, Technology, and Logistics</td>
</tr>
<tr>
<td>CoSTEM</td>
<td>Committee on Science, Technology, Engineering, and Math Education</td>
</tr>
<tr>
<td>DoD</td>
<td>Department of Defense</td>
</tr>
<tr>
<td>EHR</td>
<td>Education and Human Resources</td>
</tr>
<tr>
<td>GEMS</td>
<td>Gains in Education of Mathematics and Science</td>
</tr>
<tr>
<td>IES</td>
<td>International Education Society</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>NCES</td>
<td>National Center for Education Statistics</td>
</tr>
<tr>
<td>NDS</td>
<td>National Defense Strategy</td>
</tr>
<tr>
<td>NOAA</td>
<td>National Oceanic and Atmospheric Administration</td>
</tr>
<tr>
<td>NSF</td>
<td>National Science Foundation</td>
</tr>
<tr>
<td>NSTC</td>
<td>National Science and Technology Council</td>
</tr>
<tr>
<td>OASD/M&amp;RA</td>
<td>Office of the Assistant Secretary of Defense Manpower and Reserve Affairs</td>
</tr>
<tr>
<td>OSTP</td>
<td>Office of Science and Technology Policy</td>
</tr>
<tr>
<td>OUSD</td>
<td>Office of the Under Secretary of Defense</td>
</tr>
<tr>
<td>S&amp;T</td>
<td>science and technology</td>
</tr>
<tr>
<td>SEC</td>
<td>Science Education Center</td>
</tr>
<tr>
<td>SMART</td>
<td>Science, Mathematics, and Research for Transformation</td>
</tr>
<tr>
<td>STEM</td>
<td>science, technology, engineering, and mathematics</td>
</tr>
<tr>
<td>STPI</td>
<td>Science and Technology Policy Institute for Defense Analyses</td>
</tr>
</tbody>
</table>
Summary of DoD STEM Workshop on Evaluation

This report discusses a workshop that was held to bring together participants from multiple DoD agencies, Federal agencies, academic institutions, and other organizations to provide a high-level overview of science, technology, engineering, and mathematics (STEM) priorities across the DoD and address effective program evaluation.

15. SUBJECT TERMS
Education and Training; Evaluation; science, technology, engineering, and mathematics (STEM)

16. SECURITY CLASSIFICATION OF:
a. REPORT Uncl.  
b. ABSTRACT Uncl.  
c. THIS PAGE Uncl.

17. LIMITATION OF ABSTRACT
18. NUMBER OF PAGES
SAR 43

19a. NAME OF RESPONSIBLE PERSON
Lopez, Louie R.

19b. TELEPHONE NUMBER (include area code)
571-372-6715