# Measuring the Effectiveness of U.S. Government Security Awareness Programs: A Mixed-Methods Study

Jody L. Jacobs, Julie M. Haney, and Susanne M. Furman
*National Institute of Standards and Technology*

## Abstract

The goal of organizational security awareness programs is to positively influence employee security behaviors. However, organizations may struggle to determine program effectiveness, often relying on training policy compliance metrics (training completion rates) rather than measuring actual impact. Few studies have begun to discover approaches and challenges to measuring security awareness program effectiveness, particularly within compliance-focused sectors such as the U.S. government. To address this gap, we conducted a mixed-methods research study that leveraged both focus group and survey methodologies focused on U.S. government organizations. We discovered that organizations do indeed place emphasis on compliance metrics and are challenged in determining other ways to gauge success. Our results can inform guidance and other initiatives to aid organizations in measuring the effectiveness of their security awareness programs. While the research focused on the U.S. government, our findings may also have implications for other sectors and countries.

## 1 Introduction

The goal of security awareness training programs is to help employees recognize and appropriately respond to security issues, improving the overall security posture of organizations [14]. Various public and private industry sectors require or recommend annual security awareness training. For example, the Federal Information Security Modernization Act of 2014 (FISMA) [1] - the cybersecurity law for U.S. government organizations - mandates the implementation of security awareness training for all government employees and contractors. The European Union's General Data Protection Regulation requires organizations to provide similar awareness training [4].

Organizations may collect metrics about their security awareness programs to satisfy mandatory reporting requirements, justify resources, or demonstrate overall program success. The success of security awareness programs is often measured by the number of organizational employees completing the training, i.e., compliance to the training mandates. However, these compliance metrics may not indicate whether employee security behaviors and attitudes have been positively changed [2]. Indeed, prior literature and industry surveys have revealed that security awareness programs often fall short in changing behaviors, in part because they struggle with how to measure program impact [3, 5, 10]. Without insight into impact, security awareness programs may not be able to identify improvements necessary for facilitating behavior change while meeting employee and organizational needs.

Few studies have begun to discover approaches and challenges to measuring the effectiveness of organizational security awareness programs, particularly within compliance-focused sectors. To address this gap, we conducted mixed-methods research involving U.S. national government (federal) professionals who implement or oversee security awareness programs. Focus groups with 29 total individuals informed the development of a survey completed by 96 participants. While the research looked at multiple aspects of government security awareness programs, this paper focuses on a subset of research questions about measuring program effectiveness:

**RQ1:** How do organizations determine the effectiveness of their security awareness programs?

**RQ2:** Which types of effectiveness data do managers who oversee security awareness programs find most valuable?

**RQ3:** What are the challenges organizations face when trying to measure effectiveness?


*Workshop on Security Information Workers (WSIW)* 2022.
August 7, 2022, Boston, MA, United States.

Our study makes several contributions. We provide unique insights into how government security awareness programs approach and struggle with measuring effectiveness. This understanding can serve as a resource for security awareness professionals, organizational decision makers, and policy makers in their efforts to improve and advocate for security awareness programs. Results are informing the development of a formal publication to guide government organizations in building effective security awareness programs. While our study is focused on the U.S. government, findings may be transferable to security awareness programs in other sectors and countries.

## 2 Related Work

Measuring program success is a critical, but challenging aspect of security awareness programs. Unfortunately, few studies provide concrete recommendations on how to assess the long-term effectiveness of security awareness programs beyond knowledge-based checks, even though knowledge is not a guarantee of behavior [5, 7]. The lack of adequate measurement may in part be due to organizations' reliance on compliance to awareness policies (e.g., FISMA). Compliance-focused organizations view training completion rates as indicators of success when the emphasis should instead be on behavior change [3, 5].

For a holistic assessment, researchers suggest that organizations should use a combination of measures of effectiveness, including: number and kinds of security incidents related to training topics, user-initiated incident reporting, phishing simulation click rates, views/engagement with security awareness materials, and feedback from stakeholders via surveys and interviews [2, 5, 13]. IT experts in small and medium businesses expressed that measurement systems should take a semi-automated approach, with a combination of metrics retrieved by automated tools/processes and gathered via employee surveys or interviews [5]. There should be an accompanying visualization component for clearly communicating metrics to stakeholders. Additionally, awareness staff should recommend countermeasures to address perceived shortfalls. The cybersecurity training institute, SANS, found that organizations that assess their own program against peers tend to have greater leadership support for security awareness training, and, therefore, more success [10]. For example, a maturity model, such as the five-level Security Awareness Maturity Model [9], can serve as a peer benchmark.

## 3 Methodology

We conducted a sequential, mixed-methods research study consisting of focus groups followed by a survey. Focus groups provided an understanding of security awareness approaches and the concepts and challenges viewed as most important by participants. These insights then informed a follow-on survey distributed to a larger population. Our institutional research protections office approved the study.

### 3.1 Focus Groups

In the first phase, we collected qualitative data via focus groups. We selected a multiple-category design [6] with participants from three categories of organizations: 1) department-level organizations (e.g., U.S. Department of Labor), 2) sub-component agencies, which are organizations under a department (e.g., Bureau of Labor Statistics under Department of Labor), and 3) independent agencies, which are not in a department (e.g., Federal Trade Commission).

Participants were federal employees from diverse government agencies who had security awareness duties or were managers or executives who oversaw the programs within their organizations. We identified participants via: recommendations from security awareness colleagues; our professional contacts; online security-focused government mailing lists; and LinkedIn and Google searches.

We conducted eight virtual focus groups with 29 total participants, representing 28 unique government organizations (one focus group had two individuals from the same organization). Each focus group had 3-5 participants and lasted 60-75 minutes. Two focus groups were with individuals working in departments, three with sub-components, and three with independent agencies. Participants provided informed consent and completed an online survey to collect demographic and organizational information. Focus group sessions were audio-recorded and transcribed.

To start data analysis, each member of the research team independently coded a subset of three transcripts (one from each category of focus group) using a preliminary code list based on the focus group questions. We added new codes as needed and met several times to discuss codes and develop a codebook. Coding continued until all transcripts were coded by two researchers, who met to discuss code application and resolve differences. The entire research team convened to discuss overarching themes identified in the data and areas of interest to include in the subsequent survey.

### 3.2 Survey

Focus group insights informed the development of an anonymous, online survey. The final survey included questions about security awareness approaches and challenges. This paper focuses on a subset of questions related to measuring program effectiveness.

Recruitment methods and participation criteria mirrored those in the focus groups. The survey was open for 18 days, with 96 survey responses in the final dataset. Survey participants represented a diverse range of organizations of different types and sizes (see Table 1 in Appendix A).

We calculated descriptive statistics of quantitative responses (presented as rounded to the nearest whole number). We also conducted inferential statistical analysis to look for potential differences among organizations of different types, program sizes, and security awareness team sizes. However, we did not find any statistically significant differences for responses related to the measurement of effectiveness. Two researchers performed coding for open-ended responses.

### 3.3 Limitations

Although we recruited participants from organizations of varying sizes and types, our participants may not represent the full range of government security awareness programs. Our investigation is also limited to the U.S. government, which may have different security awareness training policies and pressures as compared to other sectors. However, given that security awareness training is common in many sectors, our findings may be transferable, at least in part, to other organizations.

## 4 Results

Since participants had the option of skipping survey questions, we report the number of responses (n) for each survey question. Direct quotes from the focus groups and open-ended questions in the survey are included to further expand upon quantitative survey results. We attribute focus group quotes with identifiers D01-06 for participants from departments, S01-11 for sub-components, and N01-12 for independent agencies. Survey participants are indicated with Q01-96.

### 4.1 Measures of Effectiveness

We asked participants how their organizations try to measure or determine the effectiveness of their security awareness program (Fig. 1). Training completion rates (84%) and phishing simulation click rates (72%) were the most popular measures of effectiveness, followed by program audits/evaluations with 67% of participants. Less than 30% selected attendance at security awareness events, employee surveys, and online views of security awareness materials. Four percent said that they do not attempt to measure program effectiveness. Sixty-four percent use at least five different measures, and only 4% selected just one measure of effectiveness.

In the focus groups and survey, participants provided details on the methods they employed to determine the effectiveness of their security awareness programs. Training completion rates and independent audits were used to demonstrate compliance with security awareness training mandates (usually, FISMA). Event attendance and views of online materials, such as newsletters or videos, were also used as indicators of reach across the organization.

Several participants indicated that their organizations made use of informal or formal employee feedback to determine if their security awareness efforts were perceived as valuable. A focus group participant remarked:

> "For all of our virtual events and at the end of our training, we have surveys for all of the participants.

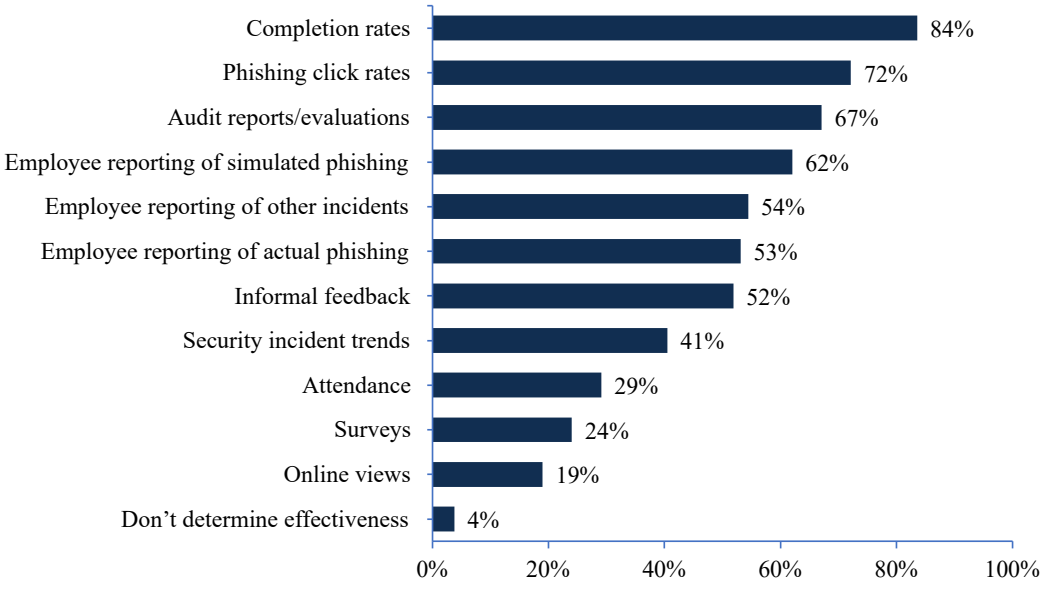


Figure 1: Measures of effectiveness (n = 79)

> And it gives them a rating scale and asks them, was the training effective? Was the content effective? Was the delivery or the presenter's delivery effective? And we use that feedback to measure our training" (D06).

Some participants looked for demonstrated employee behaviors, for example, monitoring user-generated security incident trends to determine whether certain security topics were being translated into action by the workforce. For example, a Chief Information Security Officer said:

> "There's two threat vectors that generally indicate how well the user base is aware of what's going on, particularly around improper usage and with email-based threat vectors. So, [if] those numbers are low and it makes sense, fantastic." (N06)

Organizations also used phishing simulation click rates and reporting of simulated and real-world phishing to gauge effectiveness of phishing-related training. A focus group participant said, "We analyze the metrics that come back from our phishing exercises, the reporting of those. We look at the uptick across the type of phish that we're doing,... if there's an uptick in users interacting appropriately" (N01).

## 4.2 Compliance as Indicator of Success

To determine if compliance with government mandatory training requirements (e.g., as measured by training completion rates) was regarded as the most important indicator of program success, in the survey we asked participants to rate their agreement with two statements on a five-point scale ranging from strongly disagree to strongly agree (see Fig. 2).

***"Among leadership, compliance is the most important indicator of success":*** In the first statement, participants were asked to indicate whether they agreed that their organization's leadership thought compliance was the most important indicator of security awareness program success. Over half of responding participants (56%) agreed or strongly agreed with this statement, and 22% either disagreed or strongly disagreed.

In the focus groups, several participants commented on how compliance metrics garner leadership attention, regardless of how meaningful those might be. A security awareness program lead commented:

> "We have found that, yes, management pays attention to things with compliance... We've also really found audits to be effective in helping push the cause. Now, that doesn't identify effectiveness,... but it does help increase management awareness and attention to supporting these programs" (S11).

***"I think compliance is the most important indicator of success":*** In the second statement, participants were asked to rate their agreement related to their own opinion on compliance being the most important indicator of program success. As compared to the leadership perspective, fewer (47%) agreed or strongly agreed with this statement and more (28%) disagreed or strongly disagreed.

Despite almost half of survey participants believing compliance is the most important indicator of success, many participants in both the focus groups and survey voiced a concern that compliance metrics in the form of training completion rates, although required, do not demonstrate long-term attitude or behavior change, which should be the real goals of security awareness training: "Completion of training is one statistic, but that doesn't really tell you whether anything's sunk in. It tells you that they got through the course" (N11).

## 4.3 Manager Preferences

We asked participants an open-ended question about what data would help demonstrate the effectiveness of the program to managers. Twenty-nine participants, most of whom were managers themselves, answered this question. Table 2 in Appendix B shows the types of data and example responses provided by participants.

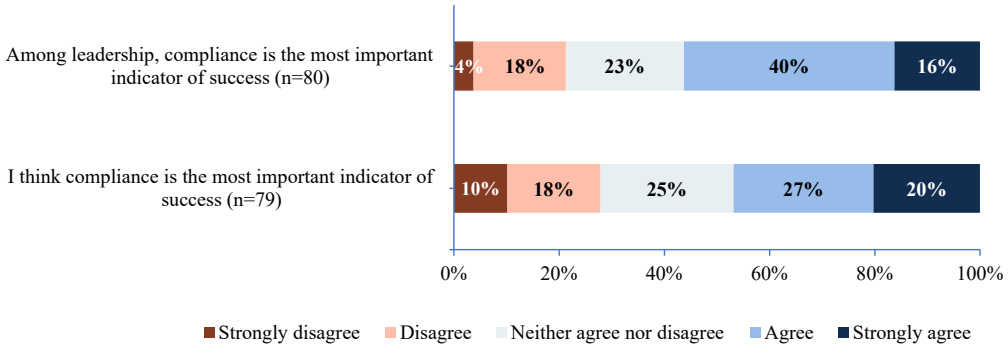


Figure 2: Agreement that compliance is the most important indicator of success

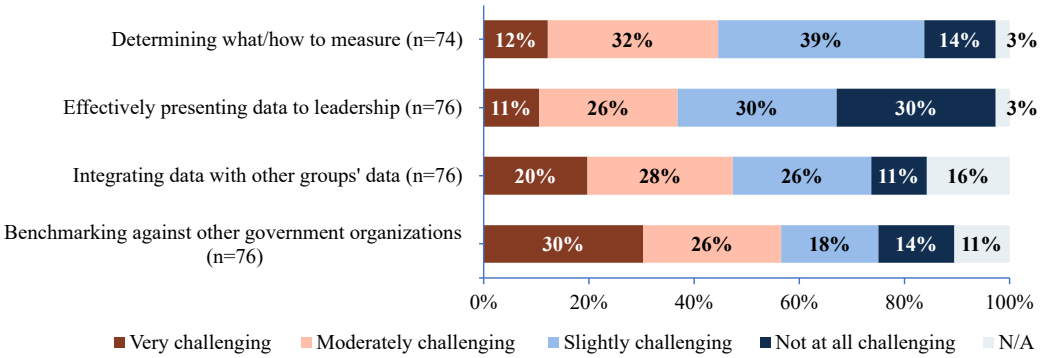


Figure 3: Challenges determining program effectiveness

Security incidents were most frequently mentioned as valuable (59% of those responding to this question). However, in the previous question on measures of effectiveness, only 41% said that their program uses security incident data, possibly demonstrating a gap in current measures. Phishing data (31%), training completion rates (24%), employee feedback (21%), and other demonstrations of employee behaviors (21%) were among other more frequently-mentioned data types.

## 4.4 Challenges

We asked participants to rate challenges experienced by their programs related to determining program success on a five-point scale ranging from "very challenging" to "not at all challenging" with a "does not apply" (N/A) option (Fig. 3). The remainder of this section provides details on survey results for each challenge and includes example supporting quotes from focus group and survey participants.

***"Determining what and how to measure":*** Forty-four percent of survey participants rated determining what to measure and how to measure program effectiveness as very or moderately challenging. Only 14% rated it not at all challenging. Although most programs make at least some attempt to determine success, almost half of focus group participants expressed uncertainty about how to gauge effectiveness. A program lead remarked, "How do we determine whether or not it is effective?...How are we making a difference when we educate our workforce?" (N04)

Participants expressed a desire for more government guidelines and standards on how to measure program effectiveness, including what variables to measure and how to interpret training metrics. For example, a participant desired "something standard that all the departments and agencies could actually end up measuring...to try and really determine whether or not the programs that are out there are effective or what parts need to actually be focused on" (S01).

***"Effectively presenting data to leadership":*** Presenting program data to leadership in an effective way was rated very or moderately challenging by 37% of survey participants. Thirty percent found it to be not challenging at all. One focus group participant expressed frustration with not being able to convince her leadership to help solve challenges faced by the security awareness team: "I have no idea how to solve the issues and challenges as, even though I have expressed challenges to the Department, it appears they all fall on deaf ears." (S09) Other participants recommended developing a robust plan to garner leadership support:

> "Write up some type of training and awareness program plan so that you can document what it is that you want the program to do and how you want it to work and all of the players that would be involved so that you can brief senior leadership on that. Because if you don't have their buy-in, then your program is probably not going to go anywhere." (D02)

***"Integrating/correlating security awareness data with data collected by other groups in my organization":*** Being able to bring together data from multiple groups to inform the security awareness program was rated as very or moderately challenging by 48% of survey participants. Only 11% said they were not at all challenged with this. Focus group participants commented on how their organizations were not currently connecting security awareness data with security incident data. A program lead said, "Ideally, you'd be able to track the incidents and see based on your security awareness and training and if your incidents are going down. We are not doing that, probably due to lack of resources" (S06).

***"Benchmarking my organization against other federal organizations":*** Over half (56%) of survey participants rated benchmarking (comparing) their organization's security awareness program against programs in other government organizations to be challenging. Several participants expressed a desire to have more government-specific information as a comparison point:

> "With our phishing exercise results, I would love to have... a standard way of looking at our agency or across agencies or across departments. We could judge apples to apples to know where we are, how we stand up to someone else, and where we could focus our training." (S08)

## 5 Discussion

Through focus groups and a survey, we explored the approaches and challenges of U.S. government organizations in measuring security awareness program effectiveness. In this section, we discuss our observations about the tension between compliance and impact as well as offering suggestions on how organizations can more effectively measure program impact and be supported in doing so.

### 5.1 Compliance vs. Impact

U.S. government organizations are required to have an awareness training component as part of FISMA [1]. Therefore, it was not surprising that training completion rates were the most common measure of effectiveness in our survey and compliance was rated by almost half of participants as being the most important indicator of success. The fact that compliance is valued as most important is concerning because research shows trainings that emphasize compliance may not translate to improved employee security behaviors [2, 3, 10].

When we asked managers what data would be most helpful to demonstrate the effectiveness of the security awareness program, they most frequently listed behavioral data (like phishing click rates, user security incidents). This was in contrast to the 56% of participants who believed that their management thinks compliance is the most important indicator of success. Furthermore, as also indicated in the survey, there were limitations in current programs for obtaining these behavioral measures.

### 5.2 Supporting Organizations

We found that government organizations struggle with knowing what and how to measure effectiveness, with a lack of standardization across organizations. The following are suggestions for how organizations can better be supported.

**Develop guidance and share lessons learned.** Those government organizations that develop security awareness guidance and policies – e.g., National Institute of Standards and Technology (NIST) and the Office of Management and Budget – could offer concrete advice on deliberate planning of how to implement measures of effectiveness, standardize measures, and provide examples of metrics to help organizations measure effectiveness. Guidance could also include how to correlate data from multiple sources (viewed as challenging by almost half of survey participants) and examples of how to effectively present data to leadership, e.g., using visualizations and ensuring data is contextually specific [12]. Also helpful would be the encouragement of security awareness professionals to utilize forums for sharing lessons learned and asking questions about measuring effectiveness, e.g., the Federal Information Security Educator's forum [8] and the SANS Security Awareness online community [11]. An upcoming revision of an older NIST security awareness publication [14] entitled "Building a Cybersecurity and Privacy Awareness and Training Program" will incorporate many of these suggestions.

**Provide benchmarking information.** Given that over half of our participants rated benchmarking as challenging, oversight organizations could also aggregate and share government-specific data to allow comparisons across programs. Maturity models like the SANS security awareness maturity model [9] can be another possible tool for organizations to gauge how their programs are doing.

**Refocus reporting guidance.** Training reporting guidance (e.g., for FISMA) currently focused on completion rates could be expanded to include reporting of metrics that emphasize evidence of security outcomes. In this case, intended outcomes are the impacts on employees' attitudes and behaviors (example measures are in the next paragraph).

**Collect data from multiple sources.** For a holistic perspective, organizations should not rely on only one metric. Rather, they can leverage and combine a variety of different types of metrics – both quantitative and qualitative – while ensuring that these are relevant to organizational decision makers [2, 5, 13]. In addition to who and how many took training, programs could look at demonstrations of employee behaviors (e.g., phishing click rates and reporting, use of secure authentication mechanisms, user-generated security incidents, security policy violations) and which types of employees or organizational groups seem to have the most issues. It is also valuable to involve employees as active contributors to the program by collecting feedback from employees about what is working or not working for them, e.g., via anonymous surveys and focus groups. Ultimately, measures should be part of an iterative feedback loop to continually identify areas of concern, refocus, and improve security awareness initiatives.

**Automate metrics collection.** Deliberate planning of *what* measures to collect should be followed by deciding *how* to collect those measures. For efficiency and consistency, quantitative metrics can be automated as much as possible [12]. For example, organizations can leverage existing technology, such as learning management systems, automatic phishing reporting buttons on email clients, or security operations data queries.

## Disclaimer

Certain commercial companies are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the companies or products identified are necessarily the best available for the purpose.

## References


[1] 113th Congress. Federal information security modernization act of 2014, Pub. L. 113-283, 128 Stat. 3073. https://www.govinfo.gov/app/details/PLAW-113publ283, 2014.

[2] Moneer Alshaikh, Sean B. Maynard, Atif Ahmad, and Shanton Chang. An exploratory study of current information security training and awareness practices in organizations. In *Hawaii International Conference on System Sciences*, pages 5085–5094, 2018.

[3] Maria Bada, M. Angela Sasse, and Jason RC Nurse. Cyber security awareness campaigns: Why do they fail to change behaviour? https://arxiv.org/ftp/arxiv/papers/1901/1901.02672.pdf, 2019.

[4] European Union. General data protection regulation. https://gdpr.eu/, 2016.

[5] Tobias Fertig, Andreas Erwin Schütz, and Kristin Weber. Current issues of metrics for information security awareness. In *European Conference on Information Systems*, 2020.

[6] Richard A. Krueger and Mary Anne Casey. *Focus Groups: A Practical Guide for Applied Research*. Sage, 2015.

[7] Khangwelo Muronga, Marlein Herselman, Adele Botha, and Adële Da Veiga. An analysis of assessment approaches and maturity scales used for evaluation of information security and cybersecurity user awareness and training programs: A scoping review. In *2019 Conference on Next Generation Computing Applications (NextComp)*, pages 1–6, 2019.

[8] National Institute of Standards and Technology. FISSEA – Federal Information Security Educators. https://csrc.nist.gov/projects/fissea, 2022.

[9] SANS. Security awareness maturity model. https://www.sans.org/security-awareness-training/blog/security-awareness-maturity-model-kit, 2018.

[10] SANS. 2021 SANS security awareness report: Managing human cyber risk. https://www.sans.org/security-awareness-training/resources/reports/sareport-2021/, 2021.

[11] SANS. SANS security awareness resources. https://www.sans.org/security-awareness-training/resources/, 2022.

[12] Lance Spitzner. Security awareness metrics. https://www.sans.org/blog/security-awareness-metrics/, 2019.

[13] Lance Spitzner. Security awareness metrics – what to measure and how. https://www.sans.org/blog/security-awareness-metrics-what-to-measure-and-how/, 2021.

[14] Mark Wilson and Joan Hash. NIST Special Publication 800-50 - Building an information technology security awareness program. https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-50.pdf, 2003.


## Appendix A: Represented Organizations

Participants indicated their organizations' type, size (number of government employees), and number of people covered by the organization's security awareness program (government employees and contractors).

Table 1: Organizations represented in the study

| | | **Focus Groups (n=28)*** | **Survey (n=96)** |
|---|---|---|---|
| *Organization type* | Independent | 42.9% | 35.4% |
| | Department | 21.4% | 32.3% |
| | Sub-component | 35.7% | 31.3% |
| *Organization size* | Less than 1,000 | 7.1% | 17.7% |
| | 1,000-4,999 | 32.1% | 29.2% |
| | 5,000-9,999 | 10.7% | 10.4% |
| | 10,000-29,999 | 17.9% | 14.6% |
| | 30,000+ | 32.2% | 25.0% |
| | Don't know | 0% | 3.1% |
| *Security awareness program size* | Less than 1,000 | 0% | 22.1% |
| | 1,000-4,999 | 25% | 25.3% |
| | 5,000-9,999 | 7.1% | 7.4% |
| | 10,000-29,999 | 21.4% | 18.9% |
| | 30,000+ | 12.8% | 24.2% |
| | Don't know | 3.6% | 2.1% |

* There were 28 unique organizations represented in the focus groups, with 2 participants from the same organization.

## Appendix B: Manager Preferences for Demonstrating Effectiveness

Table 2: Manager perspective - Data demonstrating security awareness program value (n = 29)

| **Type of Data** | **Example Responses** | **Number of Participants** |
|---|---|---|
| Security incidents | "We also need incidents more granularly analyzed and categorized as to the types of human actions/inactions that contributed, and who, so we can adjust both general training and targeted follow-up training with individuals (i.e. organizational and individual needs assessment)." (Q43) | 17 |
| Phishing data | "Effectiveness is generally measured by phishing reporting to the security team or phishing clicks during a phishing exercise." (Q74) | 9 |
| Completion rates | "Metrics for timely completion of training" (Q38) | 7 |
| Employee/user feedback | "We also review feedback of the training." (Q39)<br>"Surveys" (Q88) | 6 |
| Other employee demonstrations of behavior | "Ability to recognize areas of concern" (Q93)<br>"Adhering to the rules of behavior" (Q39) | 6 |
| Data relationships | "The data that would be most beneficial to demonstrate the value and effectiveness of the security awareness program would be the annual CSAT [cybersecurity awareness training], IT Professional/Role Based Training, and Phishing Click data graphed with the Network Monitoring data and Helpdesk reporting data on attempted non-approved access, Phishing and Spam email blocks, and other similar type data to show and compare between user knowledge and actions." (Q17) | 4 |
| Training topics | "Categories of questions pertaining to each area of operations (e.g., HR, Legal, Program & Project operations, Scientific & Engineering groups, IT specialties, Business Intelligence & Decision management, etc). Topical areas help to identify the practical application of cybersecurity across the organization and in each phase of lifecycle management/operations." (Q38) | 4 |
| Employee reporting | "The number of staff who actually recognized an incident, report them, and follow recommended practices. If staff don't do these basic things then they have not learned and the program is not successful." (Q30) | 3 |
| Participation | "Event attendance" (Q24) | 3 |
| External data | "Other federal data on compliance with training mandates" (Q41)<br>"peer agency metrics" (Q83) | 3 |
| Knowledge testing | "Exam scores, number of times a course is repeated, most likely failed questions, most often passed questions" (Q38) | 3 |