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

Jody L. Jacobs[0000−0002−6433−884X], Julie M. Haney[0000−0002−6017−9693], and Susanne M. Furman[0000−0002−7013−6603]

National Institute of Standards and Technology, Gaithersburg MD 20899, USA
{jody.jacobs, julie.haney, susanne.furman}@nist.gov
https://csrc.nist.gov/usable-cybersecurity

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 (e.g., training completion rates) rather than measuring actual impact. Few studies have begun to discover approaches and challenges to measuring security awareness program effectiveness within compliance-focused sectors such as the United States (U.S.) government. To address this gap, we conducted a mixed-methods research study that leveraged both focus group and survey methodologies centered 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.

Keywords: security awareness · training · government · effectiveness · metrics · mixed-methods.

1 Introduction

The goal of organizational security awareness programs is to help employees recognize and appropriately respond to security issues, improving the overall security posture of organizations [28]. Various public and private industry sectors require or recommend annual security awareness training. For example, the Federal Information Security Modernization Act (FISMA) - a security law for U.S. Government organizations - mandates the implementation of security awareness training for all employees [2].

Organizations collect metrics about their security awareness programs to satisfy mandatory training requirements, show return on investment, or demonstrate overall program success and value to management [5,14]. The success of security awareness programs is often measured by the number of organizational employees completing or attending the training (i.e., compliance to the training mandates) [5,16]. However, these compliance metrics tell little of how employee

security behaviors and attitudes have been positively changed [3]. 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 [4,9,22,24]. Without insight into impact, security awareness programs may not be able to identify and plan for improvements necessary for facilitating behavior change and adjusting to ever-changing threats and organizational needs [5].

Few studies have begun to discover the approaches and challenges to measuring the effectiveness of organizational security awareness programs within compliance-focused sectors like the government [17,21]. To address this gap, we conducted mixed-methods research involving U.S. Government (federal) professionals who implement or oversee security awareness programs. Focus groups with 29 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 (RQs) about measuring program effectiveness:

RQ1: How do U.S. Government organizations determine the effectiveness of their security awareness programs?
RQ2: How do government security awareness teams use program effectiveness data?
RQ3: Which types of effectiveness data do managers find most valuable?
RQ4: What are the challenges government organizations face when trying to measure effectiveness?

Our study makes several contributions. We provide new insights into how security awareness programs approach and struggle with measuring effectiveness within a yet-to-be-explored context (the U.S. Government). This understanding can inform government security awareness professionals, organizational decision makers, and policy makers in their efforts to improve security awareness programs. Results are also contributing to the development of a publication to guide organizations in building effective security awareness programs [18]. While our study is U.S. government-focused, findings may be transferable to other sectors and countries.

2 Background and Related Work

To better contextualize our study results, we provide a summary of prior literature related to measuring the effectiveness of security awareness programs and background information on security awareness mandates.

2.1 Security Awareness Mandates

Security awareness programs are meant to provide employees with an understanding of security risks and the knowledge and tools to help them take appropriate action, with a goal of achieving long-term behavior change [28]. The

cornerstone of security awareness programs is awareness training, most often conducted online and annually. U.S. Government agencies are mandated to conduct this annual training for all employees and contractors in accordance with several directives, including Office of Management and Budget Circular A-130 *Managing Information as a Strategic Resource* [20] and FISMA [2]. Beyond federal organizations, some U.S. state governments have also adopted security awareness training as a part of their information security program. For example, the Massachusetts data security law requires ongoing training focused on internal and external risks to data records containing personal information [27]. Other countries, such as Canada [10], also have training directives.

Organizations in the private sector may also be subject to security awareness training mandates. For example, organizations in the healthcare sector are required to conduct training under the Health Insurance Portability and Accountability Act [7], and the financial sector must adhere to the Gramm-Leach-Bliley Act, which requires similar training to ensure the protection of sensitive client and financial information [1].

Beyond annual, mandated training, programs – though not required – may integrate other security-related activities and communications throughout the year, including newsletters, emails, speaker events, posters, and even novel approaches, (e.g., virtual reality and escape rooms) [11]. These additional activities intend to reinforce learning and dynamically to address new security threats, policies, and processes as they arise.

2.2 Evaluating Security Awareness Programs

Measuring success is a critical, but challenging aspect of security awareness programs, with many organizations failing to adequately gauge program effectiveness [4,9]. In fact, in an industry survey of 600 organizations, less than half reported that their organizations attempt to measure the effectiveness of their awareness programs [16]. This shortfall may in part be due to reliance on metrics focused on compliance to awareness training policies (e.g., FISMA) as indicators of success [16]. However, compliance metrics fail to capture overall program impact (i.e., employee behavior change) and ignore the influence of additional awareness efforts throughout the year.

Several research and industry groups developed frameworks for measuring the effectiveness of security awareness programs. Manifavas et al. developed a tool to automate and formalize the deployment and maintenance of security awareness assessment, including metrics to measure changes in workforce knowledge, attitude, and behavior [14]. The European Network and Information Security Agency (ENISA) defined four categories of measures for security awareness evaluation: process improvement, attack resistance, efficiency and effectiveness, and internal protections [8]. The guidelines further recommended that organizations continually measure and monitor program performance and automate metrics gathering as much as possible. Rantos et al. developed a methodology for assessing the effectiveness of organizational security awareness programs [22]. They

identified two major issues that must be considered when measuring effectiveness: 1) whether the information has reached the target audience (e.g., if and how information was delivered) and 2) whether the information was absorbed by the target (e.g., if learning and behavior change has been achieved). In their survey research, the security training institute, SANS, found that organizations that assess their programs against peers tend to have greater leadership support for security awareness training, and, therefore, more success [24]. To provide this peer benchmark, they developed the five-level Security Awareness Maturity Model [23].

In a more recent effort, Chaudhury et al. conducted a systematic literature review towards defining metrics for measuring the success of a security awareness program [5]. The resultant metrics framework consisted of four overarching categories of indicators measured by quantitative (objective) data:

- ***Impact indicators*** measure changes in security knowledge, attitude, and behavior and can be measured by quantitative surveys, web-based tests, simulated attacks, or analysis of passive data (e.g., audits, risk assessments, security incidents).
- ***Sustainability indicators*** measure the value-added and impact on organizational policies and regulatory frameworks and can be assessed via changes in program funding and resources, cost-benefit analysis of the program, and percentage of awareness processes in organizational policies and processes.
- ***Accessibility indicators*** measure topic relevance, quality of training materials, and the reachability and usability of awareness dissemination channels. Indicators can be collected with quantitative surveys and analysis of passive data (e.g., attendance logs and training material hit counts).
- ***Monitoring indicators*** gauge the workforce's interest and participation in the security awareness program and leadership support. These indicators can be collected via quantitative surveys and analysis of passive data (e.g., attendance logs, training material hit counts).

The value of quantitative versus qualitative data to help measure effectiveness has been a topic of debate. Some argue for the use of only quantitative metrics (e.g., quantitative surveys, percentages of employees performing an action, analysis of incidents), saying that these are preferred since the data are more objective, repeatable, and can provide benchmarks for future evaluations [5,14]. However, others (e.g., [8,22]) suggest collecting a combination of quantitative and qualitative data. Qualitative measures (e.g., observations, detailed reports from employees, open-ended feedback forms) can be used to gauge audience satisfaction with the program, obtain ideas for improvement, and provide context and root-cause analysis to quantitative data.

Our study sought to position these prior research findings and frameworks within a new context: the U.S. Government. We also wished to gather data from the perspective of those working in security awareness programs, an approach that is in contrast to the majority of prior studies aimed at measuring the awareness levels of users [17,21].

Table 1. Focus group composition

Focus Group	# Participants	# Unique Organizations
Department #1	3	3
Department #2	3	3
Sub-component #1	3	3
Sub-component #2	5	4
Sub-component #3	3	3
Independent #1	4	4
Independent #2	4	4
Independent #3	4	4
Total:	**29**	**28**

3 Methodology

From December 2020 - July 2021, we conducted exploratory, sequential mixed-methods research consisting of focus groups followed by a survey. The National Institute of Standards and Technology (NIST) Research Protections Office approved the study. Focus groups provided an understanding of security awareness approaches and the concepts and challenges viewed as most important by participants. These insights informed a follow-on survey distributed to a larger population.

3.1 Focus Groups

We first collected qualitative data via focus groups. We selected a multiple-category design [13] with participants from three categories of organizations: 1) department-level organizations (e.g., U.S. Department of Labor[1]), 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). In the Executive Branch of the U.S. Government, there are 15 departments, over 200 sub-components, and just over 100 independent agencies. Participants were federal employees 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; security-focused government online mailing lists; and internet searches.

We conducted eight virtual focus groups with 29 total participants, representing 28 unique government organizations. Table 1 shows the composition of each focus group. Participants provided informed consent and completed an online survey to collect demographic and organizational information. Focus groups lasted 60-75 minutes and were audio-recorded and transcribed.

[1] Organization names are for illustrative purposes only and do not signify the organizations' participation in the study.

Following an analysis methodology informed by Grounded Theory [6], 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 remaining transcripts were coded by two researchers, who met to discuss code application and resolve differences. In accordance with the recommendation of qualitative methodologists [15], we focused not just on agreement but also on how and why disagreements in coding arose and the insights afforded by subsequent discussions. When disagreement occurred, we discussed as a group to reach consensus. In rare cases where agreement could not be reached, the primary coder made the final decision. 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. Table 2 shows the organizations represented in both the focus groups and the survey. As indicated in the table, participants reported their organizations' type and size (number of government employees), the number of people (government and non-government contractors) covered by the organization's security awareness program, and the number of individuals tasked with implementing the security awareness program (team size). We calculated descriptive statistics of quantitative responses. We also calculated inferential statistics to look for potential differences among organizations of different types, program sizes, and security awareness team sizes (Kruskal Walls H Test for ordinal dependent variables and Chi-square tests for categorical dependent variables). We only report significant results. For open-ended responses, two researchers performed qualitative data coding similar to the method employed for the focus group data.

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. Similar studies with other populations would be valuable.

Table 2. Represented organizations

		Focus Groups (n=28)	**Survey (n=96)**
Type	Independent	42.9%	35.4%
	Department	21.4%	32.3%
	Sub-component	35.7%	31.3%
*Size**	Less than 1,000	7.1%	17.7%
	1,000-4,999	32.1%	29.2%
	5,000-29,999	28.6%	25.0%
	30,000+	32.2%	25.0%
	Don't know	0%	3.1%
*Program size***	Less than 1,000	0%	22.1%
	1,000-4,999	25%	25.3%
	5,000-29,999	28.5%	26.3%
	30,000+	12.8%	24.2%
	Don't know	3.6%	2.1%
Team size	1 - 2	25%	33.8%
	3 - 4	53.6%	29.7%
	6 - 10	10.7%	14.9%
	11+	10.7%	21.6%

*Size = number of government employees. **Program size = number of government and contract employees covered by the security awareness program.

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 the effectiveness of their security awareness program. Response frequencies are shown in Fig. 1. Sixty-four percent used at least five different measures, and only 4% selected just one measure of effectiveness or did not measure effectiveness. Indicators of compliance to training mandates (training completion rates and audit reports) were common across both survey and focus group participants, with completion rates being the most-selected measure in the survey (84%). In the survey, organizations also frequently utilized phishing simulation click rates (72%) and reporting of simulated (62%) and real-world phishing (53%) to gauge effectiveness of phishing-related training.

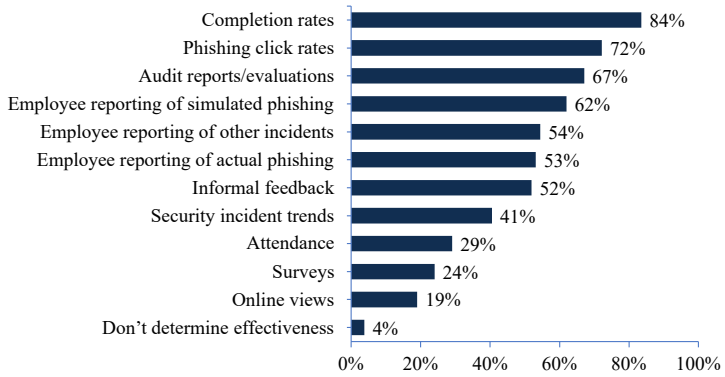

Fig. 1. Measures of effectiveness (n = 79)

Some participants looked for demonstrated employee behaviors, for example, monitoring trends in user-caused security incidents (41% in the survey) or employee incident reporting (54%) to determine whether certain security topics were being translated into action by the workforce. For example, a program lead in the focus groups said, "I interact with our SOC [security operations center] to see what types of events and incidents are being reported to see if there's any way that I can incorporate some sort of training if the incident is the result of user behavior within the agency" (N09).

Other participants made use of employee feedback to determine if their security awareness efforts were perceived as valuable: 52% of survey participants gauged success via informal feedback, and 24% used surveys. A focus group participant remarked:

> "For all of our virtual events and at the end of our training, we have surveys... It gives them a rating scale and asks them, was the training effective?... Was the delivery or the presenter's delivery effective? And we use that feedback to measure our training" (D06).

Other measures, such as event attendance and views of online materials (e.g., newsletters or videos) were also used, although by fewer survey participants (less than 30%). Several focus group participants mentioned that their organizations routinely track attendance as an indicator of reach across the organization: "We keep tally of whenever we have a speaker, we make sure that we determine all the people that are there and sort of use those as some rough stats as to success with the campaign" (S04).

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 im-

portant 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).

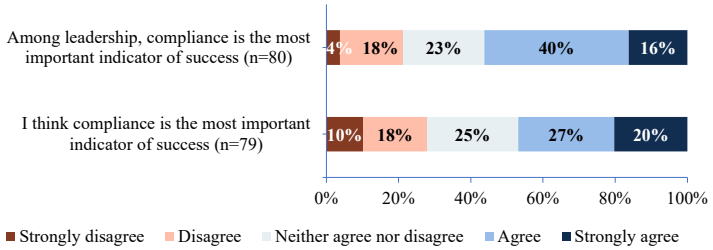

Fig. 2. Agreement that compliance is the most important indicator of success

Leadership perspective on compliance: In the first statement, participants were asked to indicate whether they believed their organization's leadership thinks compliance is 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... Now, that doesn't identify effectiveness,... but it does help increase management awareness and attention to supporting these programs" (S11).

Participant perspective on compliance: 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 with this statement and more (28%) 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: "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 Using Effectiveness Data

We asked participants how their security awareness program uses program effectiveness data (see Fig. 3). Most commonly, programs use the data to demonstrate

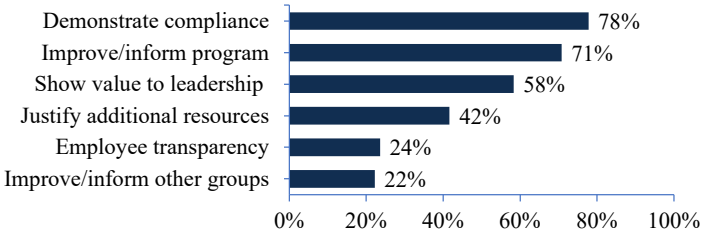

Fig. 3. How programs use effectiveness data (n = 72)

training compliance (78%) or to improve or inform the program (70%). Over half use the data to demonstrate the value of their program to leadership (58%). Less than a quarter provide the data to employees to provide transparency about the security awareness program or pass on the data to inform the efforts of other groups in the organization.

Participants provided further explanations on how they use effectiveness data. A security awareness program lead at a sub-component agency used data to inform leadership and employees within the organization: "We do have compliance metrics that we report. Management does pay attention to that, and it does heighten awareness with staff" (S11). A survey participant suggested that security awareness professionals "capture metrics to show where you started (e.g. phishing susceptibility, training rates, incident data), inform your program's strategy and tactics, and show progress" (Q43).

4.4 Manager Preferences

We asked survey participants who were managers an open-ended question about what data would help demonstrate the value and effectiveness of the program to them. Twenty-nine participants answered this question (see Table 3). 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 frequently-mentioned data.

4.5 Program Support

Perceived support in the form of direct feedback, actions, and allocated resources can be another effectiveness indicator. While we were not able to collect direct evidence of support, we were able to gauge how supported our participants felt.

Table 3. Manager perspective - Data demonstrating security awareness program value (n = 29). # indicates number of participants mentioned that data.

Type of Data	**Example Responses**	**#**
Security incidents	"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." (Q43)	17
Phishing data	"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)	6
Other demonstrations	"Adhering to the rules of behavior" (Q39)	6
Data relationships	"annual CSAT [cybersecurity awareness training], IT Professional/Role Based Training, and Phishing Click data graphed with the Network Monitoring data and Helpdesk reporting data" (Q17)	4
Training topics	"Categories of questions pertaining to each area of operations... Topical areas help to identify the practical application of cybersecurity across the organization." (Q38)	4
Employee reporting	"The number of staff who actually recognized an incident, report them, and follow recommended practices." (Q30)	3
Participation	"Event attendance" (Q24)	3
External data	"peer agency metrics" (Q83)	3
Knowledge testing	"Exam scores, number of times a course is repeated, most likely failed questions" (Q38)	3

Perceptions of Support. Participants rated their level of agreement for two statements about perceived support for the security awareness program. Figure 4 shows the agreement percentages.

Leadership support for the security awareness program: A large majority (88%) thought their leadership was supportive of their program. Only 5% disagreed or strongly disagreed. When participants were asked what advice they would provide to their colleagues, gaining leadership support was the one of the most frequently mentioned topics in both the focus groups and survey. A focus group participant commented on their leadership's support: "I would say we have good support from our management and executives. They seem to give us a lot of flexibility. If we want everyone to have a phishing exercise, they give us a little leeway to do so. If we draft a newsletter or a poster or something, they'll send it out to the user population agency-wide" (S07). However, several lamented the lack of support. A program lead said, "I don't think management would do it unless it was mandated by law... At least every few years, I have to quote the legal basis for delivering this required training" (D05). A survey participant commented, "Anything that was done in the past was personal initiative. I've done newsletters, websites, tried to get Hollywood movies regarding

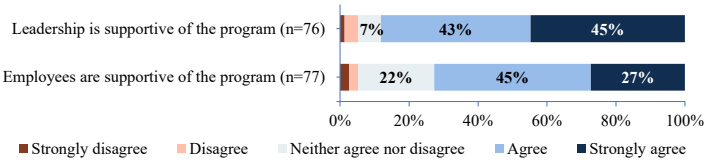

Fig. 4. Agreement that workforce supports the security awareness program

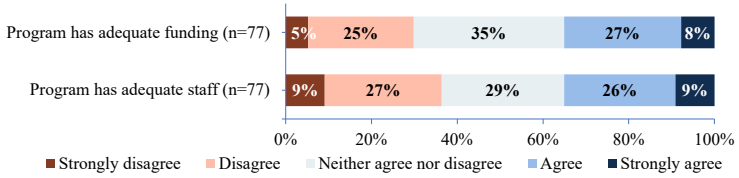

Fig. 5. Agreement about having adequate resources

security shown with discussion afterwards, tried to get a security game (I was told that was insulting). Management just didn't care" (Q34).

Employee support of the security awareness program: Seventy-two percent of survey participants agreed that employees were supportive of their program. Just 6% disagreed/strongly disagreed. When asked about successful aspects of their program, some participants commented on employee engagement and interest in the program. One said that their program was "popular with the people, encouraging engagement and behavior change" (Q56). In contrast, others remarked that employees may lack the time or motivation to engage with security awareness information or activities: "A lot of times we'll find that sometimes our users aren't engaging with the message, or they may delete it, or they don't report it the way that we want them to" (N01).

Program Resources. Survey participants indicated their level of agreement about two statements about whether the security awareness program has adequate resources in the form of funding and staff. Figure 5 shows the results.

Adequate funding for the security awareness program: Only a little over one third of participants (35%) agreed/strongly agreed that their security awareness program has adequate funding, with 30% disagreeing or strongly disagreeing. There were statistically significant differences between: very small and large teams (z = -2.445); and small and large teams (z = -2.925). For very small teams (n = 22), 32% disagreed with the statement and 23% agreed (the remainder were neutral). For small teams (n = 19), 47% disagreed/strongly disagreed and only 16% agreed/strongly agreed. In contrast, large teams (n = 12) were more likely to agree/strongly agree that they had adequate funding (67%).

When asked what could help their programs be more successful, more funding was a common response. A survey participant remarked, "Finding content is not the problem– getting funding/approval to purchase it is the problem" (Q43). A focus group participant commented, "We have a very small budget for our cybersecurity awareness program. I've seen some products in the private sector that are very slick and customizable, but they're also expensive" (S06).

Adequate staff dedicated to the security awareness program: When asked about program staffing, 35 agreed/strongly agreed that they had adequate staff, while 36% disagreed/strongly disagreed. Staff resources were closely related to perceived lack of funding. There were statistically significant differences between: very small and large teams (z = -2.198); and small and large teams (z = -2.758). While a large number of participants with very small and small teams did not think they had adequate staff (45% and 58%, respectively), only 8% (one) participant in a large organization disagreed with the statement.

In qualitative remarks, participants often discussed needing more staff to improve their programs. For example, a survey participant expressed the need for "Additional staff/SMEs [subject matter experts] to help create content other than only myself" (Q74). The fact that most security awareness team members were part-time and had other duties also contributed to the staffing shortage: "The team... who perform the security-related operations for our network, they're the same team that helps create and manage the training. So, if we have an issue or a series of issues, sometimes we may have to either delay training or make a lighter version of training" (N07). Staff turnover was also viewed as a disruption for programs: "Frequent staff turnover, including CISO and CIO positions decrease the long-term success of a program because ideas, funding and priorities change and ultimately limit program strength and growth opportunities. Meaning you can't build a great house, if you keep ripping up the foundation every year or two" (Q24).

4.6 Challenges

We asked participants to rate their programs' challenges 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. 6). 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. 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

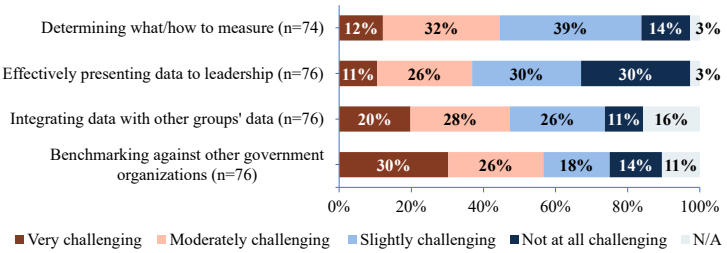

Fig. 6. Challenges determining program effectiveness

and how to interpret training metrics. For example, a participant saw the potential benefit of having "something standard that all the departments and agencies could actually end up measuring" (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. One focus group participant expressed frustration with not being able to convince their 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% rated this as not challenging at all. 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).

4.7 Perceptions of Overall Success

We asked participants to rate the overall success of their security awareness program on a four-point scale ranging from "very unsuccessful" to "very successful" (Fig. 7). Over three-quarters (77%) rated their programs as moderately or very successful. None rated their program as unsuccessful, and only 4% rated their programs as very unsuccessful. Using Kendall's Rank Correlation for ordinal data, we found a statistically significant association between the number of measures of effectiveness employed by organizations and the ratings of program success ($\tau = 0.36$, $p = 0.001$); the more measures of effectiveness used, the higher the rating of success.

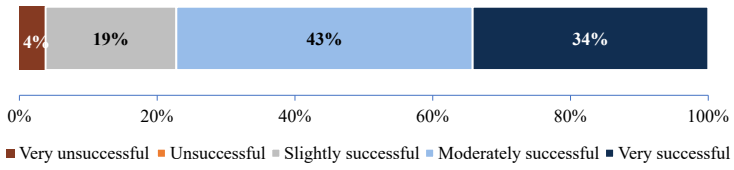

Fig. 7. Ratings of overall security awareness program success (n = 80)

During the focus groups, participants differed on the ultimate indicator of success for their security awareness program. Some emphasized compliance: "I was at 99.9% last year, which is pretty hard when you have between 38 and 45 thousand employees" (S09). However, others saw overall success as being grounded in a tangible reduction in incidents. One focus group participant remarked, "That is really the number of incidents that we end up having and tracked throughout the year and ultimately, not to be on the five o'clock news for some type of compromise or breach" (S01). Another explained:

> "It is... the elimination or... mitigation of all those threats and vulnerabilities, those incidents that have to be reported and even those that don't have to be reported. Just you want to make sure that we have smooth sailing as far as our daily operations, that there's no impact to... the service that we're supposed to provide for the federal government" (S08).

5 Discussion

In this section, grounded in our results and situated in prior research literature, we offer suggestions on how organizations can be supported in effectively measuring program impact. We also discuss areas for future work.

5.1 Development of Guidance and Standards

The following are suggestions for supporting organizations (e.g., sector oversight, standards, and training institutions) that develop security awareness guidance and policies. Organizations can also individually document their own standards and lessons learned to aid in repeatability and continuity when program staff changes.

Develop standards and share lessons learned. Most study participants said that their security awareness programs were successful. However, this raises the question of how participants know their programs are successful given their expressed challenge with determining what and how to measure and a lack of government guidance and standards, as also confirmed in prior literature focused on the private sector [17,21]. To address these challenges, guidance could include concrete advice on deliberate planning of measures of effectiveness and standardized measures. An upcoming U.S. Government document entitled "Building a Cybersecurity and Privacy Awareness and Training Program" [18], which was informed by our study, will incorporate many of these suggestions. Guidance could also include how to correlate data from multiple sources (viewed as challenging by almost half of survey participants). Additionally, because the meaning of effectiveness metrics can be contextual [5], guidance can include suggestions for how to tailor baselines of measures to the needs and risk levels of an organization.

Emphasize impact over compliance. Since U.S. Government organizations are required to conduct security awareness training, it was not surprising that training completion rates were the most common measure of effectiveness in our survey and viewed by many as being the most important indicator of success. However, as compared to findings by other researchers in the private sector [3,4,24], we observed a substantial disconnect between the emphasis on compliance and the actual purpose of security awareness: facilitating better employee security behaviors. To combat this issue, guidance documents should emphasize the importance of assessing behavioral impacts.

Provide guidance on presenting data to leadership. We also observed a disconnect in how security awareness professionals present effectiveness data to organizational leadership. While security incident trends were less commonly utilized by our surveyed organizations, participants who were also managers listed incidents as the measure of effectiveness most preferred for helping them make decisions about the security awareness program, placing less emphasis on compliance metrics. We also found a dissonance between the high levels of perceived leadership support and the high percentages saying that their programs lacked adequate funding and staff. This leaves one to wonder why leadership

support had not been translated into resources. Therefore, we see a need for guidance documents to provide examples of what kind of data is most relevant to organizational decision makers to garner both support and needed resources. Suggestions on how to effectively present that data to leadership (e.g., using visualizations and ensuring data is contextually specific [26]) can also help address participants' challenge in that area.

Facilitate benchmarking and information sharing. Given that over half of our participants rated benchmarking as challenging, oversight organizations could also aggregate and share sector-specific data to allow comparisons across programs. Also helpful will be the encouragement of security awareness professionals to utilize maturity models for benchmarking their programs (e.g., the SANS Security Awareness Maturity Model [23]) and forums for sharing experiences related to measuring effectiveness with their peers (e.g., via the Federal Information Security Educator's forum [19] and the SANS Security Awareness online community [25]).

5.2 Collect Holistic Measures of Effectiveness

Collect data from multiple sources for multiple purposes. 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 – as suggested by prior research [3,8,9,12,22,26]. Ultimately, measures should be part of an iterative feedback loop to continually identify areas of concern, refocus, and improve security awareness initiatives. Situating our findings within the metrics framework suggested by Chaudhury et al. [5], we observed an emphasis on impact and monitoring indicators, but suggest collecting the following, more comprehensive indicators:

Impact indicators: More than half of participants measured program effectiveness with phishing click rates, audit reports, and reporting of security incidents (real and simulated phishing and other incidents). In addition to these, programs could look at further demonstrations of employee behaviors, such as the use of secure authentication mechanisms, user-generated security incidents, and security policy violations.

Sustainability indicators: Sustainability indicators were only addressed tangentially in the survey in the expression of challenge programs encountered when trying to present meaningful and influential data to leadership. To remedy this current shortfall, programs could better track changes in program resourcing and influences on organizational policies.

Accessibility indicators: In addition to who and how many employees were reached by security awareness training and other communications, programs can track which types of employees or organizational groups seem to have the most security-related issues or are less likely to receive or pay attention to awareness information. These program teams could then put additional effort into reaching those populations. Accessibility indicators could also be collected via workforce surveys

(which were utilized by less than 25% of participants) to gauge topic relevance and perceived quality of materials. Furthermore, informal usability evaluations of security awareness information could be valuable in determining whether security awareness communications are properly tailored to the various workforce audiences and actionable.

Monitoring indicators: While most organizations collected training completion rates, other types of data could help assess the workforce's interest and engagement in the program. Event attendance and online views of awareness materials were less popular but could be valuable for demonstrating effort in accessing awareness information. Also helpful is the collection of both informal and formal feedback from employees about what is working or not working for them (e.g., via anonymous surveys and focus groups). Feedback from organizational leadership could also help assess impact and organizational attitudes towards 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 should be automated as much as possible [9,14,26]. For example, organizations can leverage existing technology, such as learning management systems, automatic phishing reporting buttons on email clients, or security operations data queries.

5.3 Areas for Future Research

While quantitative data can be especially helpful in identifying issues for managers, unlike Chaudhury et al. [5] and Manifavas et al. [14] who advocated for the exclusive usage of quantitative metrics, our results indicate that qualitative indicators may also be complementary as this data can expand upon quantitative indicators and get at the root cause of workforce challenges and behaviors [8,22]. Additional research is needed to develop recommendations on how programs can gather robust qualitative data and to explore how quantitative and qualitative data can be most effectively and efficiently synthesized. We also do not address potential ethical implications of the collection of effectiveness indicators, especially if used punitively against employees [5]. Additional investigation is needed to determine how data can protect the privacy of employees while still being meaningful and actionable to the organization.

6 Conclusion

Through focus groups and a survey, we provide additional evidence towards developing standards on how to evaluate security awareness programs, a current and important gap [5]. We extended prior research focused on the private sector by exploring the approaches and challenges of U.S. Government organizations in measuring security awareness program effectiveness. We found that compliance metrics were viewed as a primary indicator of program success as opposed

to impact on workforce behaviors. Organizations were particularly challenged in determining what to measure due to a lack of standards, management support, and resources across the government. Our results are informing guidance and other initiatives to aid organizations in measuring the effectiveness of their programs.

Disclaimer

Certain commercial companies or products are identified 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. 106th Congress: S.900 - Gramm-Leach-Bliley Act (1999), https://www.congress.gov/bill/106th-congress/senate-bill/900
2. 113th Congress: Federal information security modernization act of 2014, Pub. L. 113-283, 128 Stat. 3073 (2014), https://www.govinfo.gov/app/details/PLAW-113publ283
3. Alshaikh, M., Maynard, S.B., Ahmad, A., Chang, S.: An exploratory study of current information security training and awareness practices in organizations. In: 51st Hawaii International Conference on System Sciences. pp. 5085–5094 (2018)
4. Bada, M., Sasse, M.A., Nurse, J.R.: Cyber security awareness campaigns: Why do they fail to change behaviour? (2019), https://arxiv.org/ftp/arxiv/papers/1901/1901.02672.pdf
5. Chaudhary, S., Gkioulos, V., Katsikas, S.: Developing metrics to assess the effectiveness of cybersecurity awareness program. Journal of Cybersecurity **8**(1), tyac006 (2022)
6. Corbin, J., Strauss, A.L.: Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Sage, Thousand Oaks, CA, 4th edn. (2015)
7. Department of Health and Human Services: The HIPAA privacy rule (2021), https://www.hhs.gov/hipaa/for-professionals/privacy/index.html
8. European Union Agency for Cybersecurity (ENISA): The new user's guide: how to raise information security awareness (en) (2010), https://www.enisa.europa.eu/publications/archive/copy_of_new-users-guide
9. Fertig, T., Schütz, A.E., Weber, K.: Current issues of metrics for information security awareness. In: European Conference on Information Systems (2020)
10. Government of Canada: Directive on security management (2019), https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32611&section=procedure&p=H
11. Haney, J., Jacobs, J., Furman, S., Barrientos, F.: NISTIR 8420A Approaches and challenges of federal cybersecurity awareness programs (2022), https://nvlpubs.nist.gov/nistpubs/ir/2022/NIST.IR.8420A.pdf
12. Jaeger, L.: Information security awareness: literature review and integrative framework. In: 51st Hawaii International Conference on System Sciences. pp. 4703–4712 (2018)
13. Krueger, R.A., Casey, M.A.: Focus Groups: A Practical Guide for Applied Research. Sage (2015)

14. Manifavas, C., Fysarakis, K., Rantos, K., Hatzivasilis, G.: Dynamic security awareness program evaluation. In: Proceedings of the 16th International Conference on Human-Computer Interaction. pp. 258–269 (2014)
15. McDonald, N., Schoenebeck, S., Forte, A.: Reliability and inter-rater reliability in qualitative research: Norms and guidelines for cscw and hci practice. In: ACM on Human-Computer Interaction. p. 72 (2019)
16. Monahan, D.: Security awareness training: It's not just for compliance (2014), https://www.enterprisemanagement.com/research/asset-free.php/2734/pre/Report-Summary---Security-Awareness-Training:-It's-Not-Just-for-Compliance-pre
17. Muronga, K., Herselman, M., Botha, A., Veiga, A.D.: 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). pp. 1–6 (2019)
18. National Institute of Standards and Technology: Pre-draft call for comments: Building a cybersecurity and privacy awareness and training program (2021), https://csrc.nist.gov/publications/detail/sp/800-50/rev-1/draft
19. National Institute of Standards and Technology: FISSEA – Federal Information Security Educators (2022), https://csrc.nist.gov/projects/fissea
20. Office of Management and Budget: Circular a-130 managing information as a strategic resource (2106), https://www.whitehouse.gov/omb/information-for-agencies/circulars/
21. Rahim, A., Hayani, N., Hamid, S., Kia, M.L.M., Shamshirband, S., Furnell, S.: A systematic review of approaches to assessing cybersecurity awareness. Kybernetes **44**(4), 606–622 (2015)
22. Rantos, K., Fysarakis, K., Manifavas, C.: How effective is your security awareness program? an evaluation methodology. Information Security Journal: A Global Perspective **21**(6), 328–345 (2012)
23. SANS: Security awareness maturity model (2018), https://www.sans.org/security-awareness-training/blog/security-awareness-maturity-model-kit
24. SANS: 2021 SANS security awareness report: Managing human cyber risk (2021), https://www.sans.org/security-awareness-training/resources/reports/sareport-2021/
25. SANS: SANS security awareness resources (2022), https://www.sans.org/security-awareness-training/resources/
26. Spitzner, L.: Security awareness metrics – what to measure and how (2021), https://www.sans.org/blog/security-awareness-metrics-what-to-measure-and-how/
27. State of Massachusetts: Title 201 CMR 17.00 - Standards for the protection of personal information of residents of the commonwealth (2017), https://casetext.com/regulation/code-of-massachusetts-regulations/department-201-cmr-office-of-consumer-affairs-and-business-regulation/
28. Wilson, M., Hash, J.: NIST Special Publication 800-50 - Building an information technology security awareness program (2003), https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-50.pdf