JOURNAL OF CYBERSECURITY

Journal of Cybersecurity, 2022, 1–19
https://doi.org/10.1093/cybsec/tyac006
Review article

Review article

Developing metrics to assess the effectiveness of cybersecurity awareness program

Sunil Chaudhary [*†], **Vasileios Gkioulos and Sokratis Katsikas**

Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Teknologivegen 22, 2815 Gjøvik, Norway

*Correspondence address. Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Ametyst-bygget 119, Teknologivegen 22, 2815 Gjøvik, Norway. Tel: +47 93992516; Email: sunil.chaudhary@ntnu.no
†sunil.chaudhary, vasileios.gkioulos, sokratis.katsikas@ntnu.no

Received 16 February 2021; revised 5 August 2021; accepted 12 April 2022

Abstract

Cybersecurity awareness (CSA) is not just about knowing, but also transforming things learned into practice. It is a continuous process that needs to be adjusted in subsequent iterations to improve its usability as well as sustainability. This is possible only if a CSA program is reviewed and evaluated timely. Review and evaluation of an awareness program offer an insight into the program's effectiveness on the audience and organization, an invaluable piece of information for the continuous improvement of the program. Further, it provides the information required by the management and sponsor to decide on whether to invest in the program or not. Despite these advantages, there does not exist a common understanding of what factors to measure and how to measure them during the evaluation process. As a result, we have proposed evaluation metrics for the purpose. In order to do so, we performed a literature review of 32 papers mainly to extract the following data: (i) what factors did the paper measure, and (ii) how did it measure the factors? Next, we adapted the European Literacy Policy Network's four indicators (i.e. impact, sustainability, accessibility, and monitoring) for awareness evaluation to make it appropriate for evaluating a CSA program. We believe that measuring all four indicators will contribute to making the evaluation process systematic, complete, and replicable. More importantly, it will help to produce more inclusive, accurate, and usable results for the future enhancement of the program.

Key words: cybersecurity awareness, evaluation metrics, literature review, European Literacy Policy Network

Introduction

Cybersecurity is not just about technology, but it also includes the people who interact with technology and are responsible for properly implementing and operating it. Many past studies [1–3] have identified people's behaviors and actions to be responsible for most cybersecurity incidents. This could be a reason why '*people*' is considered the weakest link among the *people-process-technology* triad of cybersecurity. As the first step in handling human factors, raising the cybersecurity awareness (CSA) of people is of paramount importance.

CSA combines both gains in knowledge and positive changes in attitudes and behaviors (KAB) [4–7] that protect systems, data, and information from cyber threats. The learning achieved from CSA activities is not detailed or in-depth knowledge but only enough information to direct the attention of individuals to security issues, perceive their potential implications, and act responsibly (or make informed decisions) [6, 8, 9]. This is done by communicating the needed security information to the participants in a way so that they develop a healthy level of skepticism and motivation to act when encountering unusual situations [10]. Practically, this encompasses different dimensions, such as

- Make people realize that there are cyber risks and threats to which they are vulnerable.
- Alert people about the harmful implications of cyber threats.

- Draw people's attention to potential threat actors, ways they might or do target their victims, and critical assets they are interested in.
- Impart information about signs and mechanisms to identify cyber threats.
- Inform people about the existing security measures (tools, policies, procedures, guidelines, standards, regulations, laws, strategies, and practices) that can counteract the relevant cyber threats.
- Motivate people to timely use security measures to mitigate cyber threats.
- Make people understand the importance of cybersecurity and their obligations toward it.

CSA is a continuous process, and it is most effective when performed iteratively and focused on continuous improvement [4, 11]. It must comprehend factors like the evolving cyber threat landscape, advancements in technology, and shifts in an organization's missions and priorities to stay relevant to the target audience and optimized for the organization. But this is possible only if CSA programs are reviewed and evaluated for their effectiveness. Review and evaluation aim at evaluating the effectiveness of the undertaken iteration according to a set of pre-defined metrics, demonstrating in this way the achieved return on investment (ROI). Further, they facilitate the assessment of the program's suitability and the necessary enhancements for future iterations (e.g. weaknesses in content quality, delivery channels, and others). In other words, the results and measurements of a review and evaluation can act as critical indicators for future planning, updates, and improvement of a CSA program.

Over the last few years, several assessments, broadly categorized as output and outcome of the program [12], have emerged that can be used as indicators of the effectiveness of CSA programs. Many past studies depend on measuring and assessing one or more of the following factors to evaluate the effectiveness of a CSA program: (i) the audience interest in CSA programs generally quantified in terms of the number of participants, (ii) the reduction in the cybersecurity incidents occurred after the program, and (iii) the change in the audience's perception, knowledge, attitude, and behavior [13]. Although the first parameter is simple and demonstrates the audience's satisfaction or dissatisfaction with a CSA program, it does not convey whether the awareness program made any real difference in practice. Similarly, the second parameter cannot confirm whether the improvement in cybersecurity incidents results from a CSA program or has occurred simply because of a decrease in cyberattacks impacting the audience after implementing a better firewall and network protection. The third parameter is complex, but is the most relevant. It measures and assesses the changes in the security knowledge, attitude, and behavior of the audience.

Evaluations of security knowledge, attitude, and behavior use both subjective (e.g. ask the audiences about their experience), and objective (e.g. ask the audiences to do something) methods [14]. Measuring the knowledge and comprehension of security is conducted through, e.g. online quizzes that can reveal whether people know and understand the risk [8]. Similarly, measuring attitude is conducted through, e.g. anonymous surveys on why people take risky actions. But measuring the behavioral change is not simple and is performed mostly using indirect measurement, e.g. self-reporting and surveys. A few studies used simulated attacks (i.e. attack simulation based on the threat profile of the target audience) and system data (e.g. analysis of audit logs) to understand the audience's security knowledge, attitude, and behavior from their responses. Other ways include investigating the three key elements of the Fogg Behavior Model, i.e. motivation, ability, and trigger [15] that have to occur simultaneously for behavior change.

Although the aforementioned parameters and their evaluation techniques are relevant, there do not exist commonly agreed and understood standards on what constitutes an effective and successful CSA program [16] and measurements to evaluate its effectiveness [11], thus hindering the evaluation process. This may have happened because of the logic that different audiences have varying needs and situations for a CSA program, thus their intention for evaluation cannot be captured by metrics valid for all [17]. Ironically, this lack of metrics has become a major reason for organizations' struggle to determine and measure the effectiveness of their CSA program. Many organizations either do not make any provision to measure the effectiveness of their program or their evaluation is merely based on the program's outreach [17]. Without a proper evaluation, a mature CSA program is presumably unachievable [18] and above all, it has a high chance of failure. Incidentally, the SANS security awareness maturity model [19] has the final (highest) level "*metrics framework*" that also reinforces a robust metrics framework to track progress and measure impact as the main requirement for a mature awareness program. The model's main argument is that an awareness program can be called mature only if it has developed the capability to demonstrate its continuous improvement, ROI, and value to the organization. This capability is possible only if an awareness program has standard measurements or metrics in place. The need for suitable metrics to measure the effectiveness has become even more crucial considering the fact that most existing efforts to improve people's security behavior are failing to produce the desired impact [7].

Therefore, the main objective of this paper is to define the right metrics for the evaluation of a CSA program and the corresponding methods that provide information on how well the metrics have been met. Defining such metrics will help to reduce the ambiguity in the evaluation of a CSA program by indicating the priorities that need to be focused on. Moreover, they will help to assess the aspects of the program and identify what has been successful and what has not, as well as what has been a required improvement. In order to define the metrics, we used a systematic literature review and the European Literacy Policy Network's (ELINET) four indicators for awareness evaluation.

Through this proposition, we intend to make the evaluation process of a CSA program as inclusive, complete, and unbiased as possible and, more importantly, make it replicable so that everyone should be able to conduct the same evaluation and get similar results. We believe that this will help CSA professionals assess their implementation to get more accurate and usable findings for the future iterations of their awareness program and attain a successful CSA program.

Related Works

There are some major works that proposed methods for the evaluation of a CSA program. A survey report by the European Union Agency for Cybersecurity (ENISA) [5, 11] found that in general there are four main approaches, each with different performance indicators, used by organizations for the measurement and assessment of the effectiveness of CSA activities. We present details of these four approaches in Table 1. Most organizations use a blend of these approaches for assessment and make their decisions based on the overall picture rather than on a single measure. Along with that, the latter work [5] also mentioned that as the needs and situations of target groups differ greatly so should their evaluation metrics. Thus, it provided 71 key performance indicators (KPIs) and suggested con-

Table 1: CSA evaluation metrics [11]

Approach	Description	Performance indicator
Process improvement	Measures the effort invested to conduct a program (e.g. development, dissemination, and deployment) and has no link to the end result, i.e. whether security has improved or not. *Advantage* • Easy to define and gather. *Disadvantage* • Provides only indirect comfort.	• Counts the main security risks or technology platforms covered. • Counts staff reached. • Cost of delivery (time and expenses invested per person. • Relevancy of the awareness material (the frequency with which it is updated). • Staff feedback on the awareness impact (use survey).
Attack resistance	Measures how resistant the staff is to a potential attack. *Advantage* • Provides direct evidence of the actual state of staff awareness. • Important to impress top management/sponsor and receive support and commitment to the program. *Disadvantage* • Many attack scenarios and all of them cannot be tested. • Simulated tests can be relatively expensive to set up.	• Staff ability to recognize attacks (using survey, quiz, or computer-based test). • Staff susceptibility to falling prey to attacks (using simulated attacks).
Efficiency and effectiveness	Measures the experience of security incidents within the organization. *Advantage* • Easy and inexpensive to collect data. • Statistics are usually of interest to senior management. *Disadvantage* • Does not provide a true reflection of security awareness (low-security incidents can happen because of other reasons).	• Extent of security incidents (number and cost of security incidents), downtime (availability of systems is critical), and most severe incidents (a proportion of the total number of serious incidents) caused because of human behavior.
Internal protections	Measures secure behavior results because of awareness. *Advantage* • Provide direct evidence of staff security behaviors. *Disadvantage* • Measure is quite specific to the behavior it is measuring.	• Extent to which security is incorporated into the development and acquisition of systems (measured by the review of business cases and requirements specifications). • Extent to which data files are protected is measured by the review of malware infection as shown by anti-virus activities or statistics and measured by the report on visits to inappropriate materials or unauthorized software (from scanning tools).

sidering different layers (i.e. business layer, service layer, and operational layer) and dimensions (i.e. planning, managing, and evaluating), while identifying the evaluation metrics and KPIs for the evaluation of the CSA of an organization. Further, it recommended making use of industry-standard performance management models, such as the Balanced Scorecard or Six Sigma, to define performance targets and measurements. As a part of the overall evaluation of CSA, these reports suggested both formative and summative evaluations [20]. However, a major difficulty in using the metrics would be deciding what to measure and for whom it is measured (intended users of the outcomes) among the number of KPIs they recommend measuring, which has been left up to the evaluator to decide.

Another similar study by Manifavas et al. [21] suggested 12 quantitative metrics for the evaluation of the CSA of an organization, shown in Table 2. It showed a method to assign a weight to metrics. In addition, it proposed the cost of implementing and running the CSA program (i.e. cost-benefit analysis) as a part of the evaluation process. The effectiveness of an awareness program is determined by the weighted summation of the value of its underlying metrics and the summation of their (i.e. metrics) costs. But a major limitation of this study is that it does not provide a complete evaluation of any dissemination methods. For example, email views or poster downloads can convey only the reachability of an awareness program. With this information on hand, it will not infer, e.g. if the awareness message has been read and understood properly as well as practiced in everyday life by the audience. Not to mention, it is widely recommended to use multiple dissemination channels to fulfill the needs and preferences of diversified end-users [22, 23] and to retain the information richness of the awareness content as much as possible [24]. This also implies that evaluation has to be performed for each dissemination channel used.

Next, the study by Bitton et al. [25] proposed a framework for assessing the Information Security Awareness (ISA) of smartphone users. The framework focused on measuring the behavior of smart-

Table 2: Evaluation metrics and measuring parameters [21]

Approach	Description	Performance indicator
Surveys	Questionnaire-based survey on technical and security policy issues.	Statistical analysis of monthly survey (conducted in the different divisions of the organization) and annual survey (conducted in the whole organization).
Awareness/security day	Direct communication with employees to get their feedback.	Statistical analysis of security day attendance.
Independent observation	Silent observation of employees' security behaviors.	Statistical analysis of unsuccessful mock phishing attacks, and new threats bulletins' readership.
Audit department reports	Security awareness related incidents identified by audits should decline.	Count of security incidents caused due to employee behavior identified by the audit department.
Risk department reports	Risk identified during the previous assessment should reduce throughout time.	Count of security issues occurred due to employee behavior identified by the risk department.
Security incidents	The volume of security incidents that occurred.	Number of employees who have caused at least one security incident • due to their non-secure behavior (out of the total number of employees) • that falls within their responsibility but occurred due to their failure to identify the threat (out of the total number of employees).
Awareness sessions (workshops)	Post-session feedback from employee.	Statistical analysis of session attendance and effectiveness.
Information security website	Employee interest in the awareness program.	Statistical analysis of information security website visit.
e-Learning	Reachability of the awareness program and the employees' interest in it.	Statistical analysis of e-learning program visits, registrations, and completion.
Emails	Employees' interest in the awareness program (link can be provided for follow up information).	Statistical analysis of email views.
iNotices	Employees' interest in the awareness program (link can be provided for follow up information).	Statistical analysis of iNotice reading.
Posters	Independent observations, combined with electronic means, e.g. Quick Response (QR) code to additional resources, or Uniform Resource Locator (URL) from where the poster can be downloaded.	Statistical analysis of poster downloads.

phone users by collecting and analyzing data from three sources, which are

- a mobile device agent installed on the subject's device,
- a network traffic monitor (network traffic sent to/from the subject's devices), and
- survey data (using a security questionnaire).

The main issue in this study is the data collection methods it suggested. Data collection methods like a mobile device agent and a network traffic monitor can be controversial to use and may be considered equivalent to spying on the audiences.

Last but not least, Fertig et al. [18] performed a systematic literature review to identify metrics that are regarded for ISA and performance measurement systems (PMS) used for the assessment of ISA. They mainly found two types of metrics, which are

- Knowledge-based metrics: impacts of knowledge in attitude and behavior (KAB based [26]), evaluation of ISAs training (based on various KPIs), dissemination of knowledge (ease of sharing and using knowledge), and impacts of knowledge on individuals (improvement in the sense of responsibility for security).
- Behavior-based metrics: improvements in security behavior (for various security threats, and security practices).

In the same study, the authors also identified requirements for metrics in theory (used literature review) and practice (used interviews with experts). Further, they analyzed whether the existing metrics and PMS meet the identified requirements. Their analysis revealed that most existing metrics and PMS do not meet primarily two requirements, which are automation (computed automatically) and visualization (visualizing the results properly). However, the study does not discuss how metrics can be measured (measurement methods).

All the aforementioned works do not dismiss the value of qualitative methods in the evaluation process; however, they all emphasize the use of quantitative methods. This is mainly because it is relatively easier to express and convey the message on the benefit of CSA to the senior management using quantitative values, i.e. in a language and format they understand and are used to. Moreover, all of them primarily focus on the evaluation of CSA in organizations.

Indeed, the aforementioned works are useful; however, we believe that they disregard certain important aspects to measure that we have included and compared with our proposition in Section Metrics Development.

Research Methodology

The main objective of this research is to come up with appropriate metrics for evaluating a CSA program. To begin with, we have conducted a systematic literature review with an intent of determining and analyzing the "*factors to be measured*" and their "*measuring methods*" commonly used to evaluate the performance of a CSA program. And, to conduct the literature review, we followed the structure

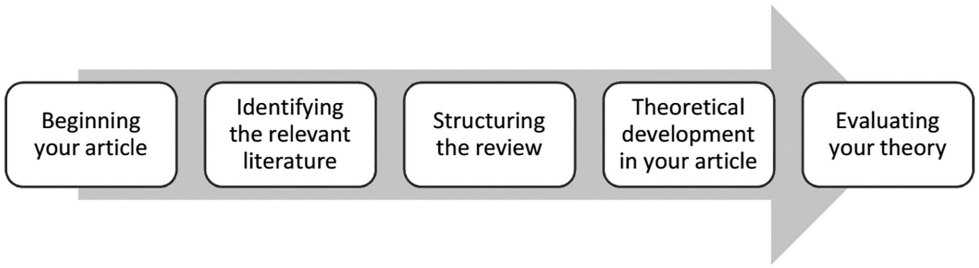

Figure 1: Structure of a literature review [27]

of a systematic literature review as suggested by Webster and Watson [27], as shown in Fig. 1. Following that, ELINET's four indicators for awareness evaluation (metrics in practice but for purposes other than cybersecurity) have been adapted for CSA purposes. For each indicator, we have provided appropriate "*factors to be measured*" and their "*measuring methods*" that should be used while evaluating the performance of a CSA program.

In order to identify relevant literature on the topic, we used search services on Google Scholar Citations and Microsoft Academic. Both are freely accessible web search engines indexing the full text or metadata of an array of scholarly materials including peer-reviewed academic journals and conference papers, books, theses and dissertations, technical reports, and other scholarly literature published in different digital libraries and databases. There are several features of these search engines that helped to accomplish the search goal more efficiently. These search engines perform search operations on the full text of documents and rank them according to their weight calculated using different important criteria [28, 29]. This feature saved lots of our time and effort that would have been required for screening relevant literature if used a search engine that indexes only the title, abstract, and keywords of a document. Further, it is a very challenging task to decide which digital libraries and databases to include in the study and know if they will result in relevant literature or not. Not to mention, many academic databases do not contain gray literature like white papers, technical reports, theses, and dissertations, which are equally useful for a systematic literature review [30, 31]. Some potential benefits of including gray literature in a review can be, e.g. this helps to reduce the possibility of publication bias, improves the comprehensiveness and timeliness of the reviews, fosters a more balanced understanding of the available evidence, and ultimately enriches the review's findings [31]. Moreover, this study does not intend to limit only to academic findings but also include industry findings, i.e. to provide a more balanced picture based on both theory and practice [18], which would be possible only by including gray literature. More importantly, performing search and screening operations in various digital libraries, university academic repositories, and others to collect a large number of relevant literature studies (including gray literature) would be exhausting. This became relatively easier and more convenient merely by searching these two search engines, as it would otherwise have required performing search operations on different databases independently.

This paper is a revised and extended version of the CyberSec4Europe project report [32]. So, we performed the first round of search operations in October 2020. With an intent to extend the report, we performed the second round of search operations in July 2021. We used "security + awareness + effectiveness," "security + awareness + success ," and "security + awareness + value," as the search keyword strings, where '+' is an "AND' operator. Before selecting this keyword string, we performed a trial with other keyword strings like "cybersecurity + awareness + effectiveness," "cyber-security + awareness + effectiveness," "cyber security + awareness + effectiveness," "information security + awareness + effectiveness," and "Internet security + awareness + effectiveness." But by using these keywords, we did not find relevant literature showing up on the top result pages. In the case of search results common to both of the search engines, a download was just made from one of them.

After a manual screening of 350 results in Google Scholar Citation and 400 results in Microsoft Academic based on their abstract and keywords, we downloaded 78 papers. The downloaded papers were thoroughly read in the second round of screening to determine how relevant the papers were to the research topic and to provide answers to the two questions "*what to measure*" and "*how to measure*" to evaluate the effectiveness of a CSA program. After the second round of screening, we selected 32 papers for the literature review comprising 19 journal papers, 12 conference and workshop papers, and a NIST Technical Series publication.

For the selection of literature, we defined inclusion and exclusion criteria as follows:

- Literature in languages other than English is excluded, since the working language for this study is English.
- No exclusion criterion was defined for the year of publication. The oldest and the most recent papers considered for the review are 2003 and 2020, respectively. Many traditional methods used for raising awareness are still relevant and in practice with necessary modifications; e.g. different organizations still use posters and leaflets [33] to raise CSA. So, restricting by the year of publication in the literature selection will lead to missing many important papers with information even now relevant and useful.
- Along with academic literature, gray literature as recommended by Kitchenham [30] and Paez [31] was also included in the review. However, only technical reports from reputable organizations also cited by many other works have been considered for the review to maintain academic integrity in the work.
- Finally, for high-quality literature as recommended by Webster and Watson [27], information like published venue (peer-reviewed journals, conferences, and workshops) and citations were used for the literature selection.

In order to structure the review, we were highly reliant on a tabular presentation style, since it is easier to present a large amount of data in an understandable form.

The theory development is based on ELINET 's four indicators for awareness evaluation, a well-established model designed for aware-

ness evaluation. We have adapted it to make it applicable to CSA. While doing so, we have considered the findings of the literature review, and some criteria for good metrics.

Data Collection and Analysis

A review of the 32 selected papers [4, 21, 26, 34–62] was performed mainly to gather "*what factors are measured or suggested to be measured*" in order to determine the effectiveness of a CSA program and "*how those factors are measured or suggested to be measured*." The collected factors and their measuring methods have been listed in Tables 3 and 4, respectively.

Factors evaluated

With a motive to evaluate the effectiveness of a CSA program, the reviewed papers measured the following factors given in Table 3, and the count of papers measuring them are shown in Fig. 2.

Knowledge, attitude, and behavior

Among the measured factors, knowledge, attitude, and behavior are found to be the most popular. Knowledge refers to a familiarity, awareness, or understanding of security policies, procedures, standards, guidelines, laws/directives/regulations, strategies, technologies/systems, and good practices. Similarly, an attitude refers to beliefs, opinions, thinking, or feelings toward security. An attitude can be positive (e.g. I am aware of my role in protecting the organization against potential cyber threats) or negative (e.g. I think cybercrime reporting is a waste of time). Finally, behavior is the way in which a person acts or conducts toward security issues, i.e. either avoiding or bringing into practice security knowledge s/he has learned. The popularity of these factors may have happened due to the influence of the KAB model (also known as the knowledge-attitude-practice-KAP model), which is widely popular in health education but now has also been increasingly adopted for CSA purposes.

The KAB model divides the changes in behavior into three successive processes that initiate with the acquisition of knowledge, followed by the generation of attitude, and finally, result in the formation of behavior or actions. The model implies that just because someone possesses security knowledge does not mean they will utilize it to good use and act appropriately. If someone has a negative attitude, their actions (behavior) will be in direct opposition to their understanding. Behavioral change comes with knowledge through attitude. However, every knowledge and attitude may not translate into action (i.e. intention–behavior gap). Many unfavorable factors, for instance, lack of suitable knowledge, adequate time and resources, and others, may inhibit translating intention into behavior.

The primary objective of a CSA program is to motivate or influence the participants in the adoption of secure online behavior [7]. This may be a reason why several studies have utilized different psychological, social, and behavioral modeling theories to study the impact of knowledge and attitude on security behaviors with a motive to improve the effectiveness of CSA [63]. Moreover, in the various classifications of CSA levels [64, 65], the highest or ultimate awareness level is also about behavioral changes, i.e. security activities and behaviors occur automatically in an individual when performing personal and professional activities. Therefore, there is no way to evaluate a CSA program's success just based on knowledge measurement but requires also including the measurements of attitude and behavior.

A major issue with the studies evaluating security knowledge, attitude, and behavior is that they develop their own measurement, which is often non-standardized (do not follow a standard process to design questionnaire and scale as well as analyze the data). They often examine only one or a few selected components of cybersecurity for the assessment. Therefore, it is recommended to use a standardized questionnaire and scale to measure knowledge, competence, attitude, and behavior [46]. Several studies have produced well-validated and standardized scales or questionnaires intended to measure security knowledge, attitude, and behaviors. Some of their examples are

- Human Aspects of Information Security Questionnaire (HAIS-Q) [66] is used to assess the vulnerability of organizational critical assets caused by the risky or risk-taking behavior of employees.
- Security Behavior Intentions Scale (SeBIS) [67] is used to measure users' self-reported adherence to computer security advice.
- SA-6 scale [68] is used to assess and compare users' security attitudes.
- Rajivan et al.'s [69] questionnaire containing a combination of skills and knowledge-based questions is used to measure security expertise in end-users.
- Hadlington's [70] combination of four scales, which are the Abbreviated impulsiveness scale (ABIS) [71], Online cognition scale (OCS) [72], Risky cybersecurity behaviors scale (RScB), and Attitudes toward cybersecurity and cybercrime in business (ATC-IB), are used to measure the human factors in cybersecurity.
- Ögǎütçü et al.'s [73] four independent scales; namely Risky Behavior Scale (RBS), Conservative Behavior Scale (CBS), Exposure to Offence Scale (EOS), and Risk Perception Scale (RPS), are used for the assessment of security-related behaviors and security awareness levels of information system (IS) users.
- Smartphone Security Behavior Scale (SSBS) [74] is used to measure the influence of mental health issues on smartphone security behavior intentions.
- Users' ISA Questionnaire (UISAQ) [75] is used to measure users' security awareness (security knowledge, belief, and behavior).

We believe that using these standard scales and questionnaires for evaluation purposes can help to get more reliable and scientifically valid results. Finally, touchability (i.e. awareness information is perceived positively) is also mainly associated with knowledge, attitude, and behavior.

Usability

Usability is the next popular factor measured to evaluate the effectiveness of a CSA program. This is measured in terms of the relevancy of topics covered, the quality of the content, and the suitability of the dissemination channels used. Reachability (i.e. the ability to reach the right audience) is also a quality of dissemination channels.

In an organization, CSA can be undertaken at different levels, e.g. individual, departmental (or business unit), and organizational [76] aligning with the individual needs, departmental (or business unit) objectives, and organizational strategic plans and goals, respectively. Organizations depending on CSA designed based on the "*one-size-fits-all*" approach completely underestimate its purpose. By doing so, the problem of cybersecurity cannot be addressed, rather it only contributes to the organization's overhead.

It is not by any means beneficial for organizations to ask their employees to mandatorily attend a CSA program on topics irrelevant to them. Essentially, CSA topics suitable for managers and executives to carry out their managerial roles and responsibilities [77],

Table 3: Factors measured by the reviewed papers to evaluate the effectiveness of a CSA program

Measured factor	Paper
Improvement in cybersecurity **behavior** resulted from participating in an awareness program. This has been expressed as follows: • Reduction in cybersecurity risky behavior. • Promotion of the best practices and compliance with safe behavior. • Positive effect on cybersecurity behavior. • Intended change in cybersecurity behavior. • Intention-in-action to change cybersecurity behavior. • Improvement in performance (i.e. change in behavior improving the effectiveness of security actions.) • Deterrent effectiveness (i.e. discourages risky actions). • Level of audience's risky behavior. • Self-responsibility to behave securely.	References [4, 26, 34–55]
Positive changes in the cybersecurity **attitude** of the audience resulted from participating in a CSA program. This has been expressed as follows: • Developed a positive attitude toward cybersecurity. • Intended change in cybersecurity attitude. • Normative belief and subjective norms toward cybersecurity. • Improvement in adaptability (i.e. opinion on how to determine and react efficiently to unsecure situations occurring unexpectedly). • Level of audience beliefs about cybersecurity. • Willingness to behave securely. • Willingness to learn about cybersecurity. • Intention in words to make positive changes in cybersecurity behavior.	References [26, 34, 35, 36–37, 40, 41, 43–50, 51, 54, 55]
Cybersecurity **knowledge and competence** gained by participating in a CSA program. This has been expressed as follows: • Cybersecurity knowledge gained. • Level of audience CSA. • Cybersecurity knowledge and competence gained. • Improvement in learnability (i.e. gain in knowledge or learn from past actions to improve the current security actions.)	References [26, 34–36, 38, 40, 43–46, 48–50, 52, 54–59]
The audience, organizer, and management/sponsor's **interest** in a CSA program. This has been expressed as follows: • Audience interest in an awareness program. • Audience interest to participate in an awareness program. • Motivation demonstrated by the organizer of an awareness program. • Manager or sponsor support and commitment to an awareness program.	References [4, 37, 40, 41, 55, 60]
Reachability of an awareness program, i.e. information has reached the right audience. This has been expressed as follows: • Awareness information reached the target audience. • Diffusion level of delivery methods.	References [21, 42, 53, 61]
Touchability of an awareness program, i.e. information is perceived positively by the right audience. This has been expressed as follows: • Awareness information touched the target audience. • The target audience absorbed the delivered information.	References [21, 61]
Value added by an awareness program, i.e. economic, or other benefits. This has been expressed as follows: • Contribution by an awareness program. • Cost-benefit analysis of an awareness program. • Cost of implementing and running an awareness program (i.e. cost-saving).	References [4, 21, 41]
Usability of topics covered, learning methods used, and awareness program organized. This has been expressed as follows: • Relevancy of awareness topics covered. • Relevancy and usefulness of awareness topics. • Relevancy or suitability of topics. • Usage of knowledge gained from awareness in practice. • Confidence gained (i.e. learned things that are useful in real life). • Preference for learning method used in an awareness program. • Preference (i.e. liked the content and delivery method used). • Satisfaction from an awareness program. • Satisfaction (i.e. learned things and enjoyed learning). • Delivery assessment. • Usability of an awareness program.	References [37, 40–42, 49–51, 60, 62]

Table 3: Continued

Measured factor	Paper
Overall feedback on an awareness program. This has been expressed as follows: • Feedback on an awareness program. • Feedback strategy. • Audience feedback.	References [4, 40, 43, 53]

cannot be of use and interest to other general employees. Likewise, awareness topics appropriate for employees with specialized roles and responsibilities, e.g. accounting, may not be of use and interest to employees in the IT department. Even within the same department, some employees may have different CSA needs than their departmental colleagues, whereas employees from different departments may have the same CSA needs. So, the evaluation must check whether the awareness topic was of use and interest to the audience and also identify the topics they really want to learn about.

The quality of CSA contents is not just about what has been expressed but also how they have been presented [78]. The same message can be framed and conveyed in a multitude of ways without changing its meaning and facts. And the way it has been conveyed largely determines whether the message will persuade the recipients or drive them away. The concern on a CSA content's presentation, therefore, should not limit just to what formats of media be used, but much more than that. Unfortunately, there is no consensus on how the quality of awareness content should be evaluated. Existing evaluations are largely based on asking the audience how informative and useful the content is, or how satisfied the audience is with the content. Such evaluations may not provide results that can disclose the aspects requiring improvement or update in the content. Therefore, we suggest a more comprehensive evaluation, where questions ask about attributes like

- Accurate, consistent, up-to-date, and complete information.
- Clear and concise presentation.
- Effective message framing.
- Convenient and doable suggestions.
- Innovative, engaging, localized, and useful message.

Finally, a CSA message is disseminated through a variety of channels, e.g. workshops, newsletters, posters, screensaver, emails, games, videos, audio, simulation, and so on. Different audience groups can have varying preferences regarding dissemination channels [79], e.g. young people may prefer dynamic channels (like games and simulations) over static ones (like posters and newsletters). Therefore, it is necessary to know how well-fit the dissemination channel was to the audience. Once again, there is not a common understanding or standard for the evaluation of dissemination channels. Certainly, there are some studies [5, 24, 80–82] that have compared various dissemination channels to gauge their advantages and limitations. Some parameters drawn from those studies, which we believe can be useful for designing evaluation questions, are

- Cost and technology (did it require any additional cost and technology to operate?).
- Operation (how easy was it to operate?).
- Work culture (did it support the users' work culture?).
- Flexibility (did it support self-paced learning?).
- Interesting and engaging (was it interesting to use and offered high engagement of users?).
- Content type (did it support preferred content types?).
- Reachability (did the information reach the right audience?).

Interest

The individual interest of the organizer, sponsor, senior management, or audience toward a CSA program is another factor widely measured to evaluate its effectiveness. In psychology, focused attention characterizes interest, increased cognitive and affective functioning, and persistent effort [83]. It has also been found to be strongly related to motivation, behavior, and outcome [84]. The organizers, sponsors and senior management interest is necessary for the sustainability of the program. Similarly, the audience's interest is essential to motivate them to participate, learn, and benefit from a CSA program. Therefore, this can be an essential factor to be measured to know the effectiveness of a CSA program.

A simple and direct way to measure interest is to ask an individual if the program interested him/her (self-reporting). But a more reliable result can be obtained by watching other indirect indicators. Indicators like voluntary participation (e.g. attendance or visits to awareness resources available online), seriousness for learning (e.g. further inquiry with a desire to learn more, or visits to additional materials), activities during physical participation, and affect or performance after participation in a CSA program (e.g. test results and changes in behavior) can be utilized to measure the interest of the audience. Next, the organizer's and sponsor's interests can be realized from the continuity of the program. Finally, in an organization, the senior management's interest can be realized from their commitment and moral support for the program, their participation in the program, and the funds allocated for the program.

Value added

Value-added, means economic (time- and cost-saving that could have suffered due to a cyberattack or repairing and reinstating normal business operations after suffering an attack) and non-economic benefits (market competitiveness, or improved customer confidence) benefits gained due to a CSA initiative. It is, once again, important for the sustainability and continuity of a CSA program. An organization performing a *cost-benefit analysis* of a CSA program [5] can provide more formal results valuable to the management. Some other indicators are lowered cyber incidents, awards received, a reputation built due to improved cybersecurity, and CSA's lessons learnings integrated into the work culture of an organization. However, such specific indicators may not be available in the case of a CSA program for civilians. In that case, a self-reporting method can be used to gather the different benefits they gained by participating in a CSA program.

Methods used for evaluation

The methods that have been utilized to evaluate each factor are presented in Table 4. Basically, these methods (shown in Fig. 3) can be classified into two types: intrusive evaluation, and non-intrusive [85].

In intrusive evaluation, the participant's normal behavior is consciously disrupted by the evaluation processes. This also means that

Table 4: Factors measured and their respective measurement methods

Measured factor	Measurement method

Behavior	**Intrusive method**
	• Questionnaire-based survey (qualitative; open-ended questions; and pre- and post-survey), face-to-face meeting, semi-structured interview, and group discussion.
	• Laboratory experiment and a non-disguised observation of participant's actions.
	• Web-based test by using vocabulary and scenario type questions.
	Non-intrusive method
	• Simulated attack and its response observation (pre- and post-attacks), e.g. count of response to the simulated phishing emails sent, or click-through rate of malicious link in an email, or download count of a malicious attachment in an email.
	• Practical system data to measure an increment in compliance with the best security practices, or a reduction in risky behaviors.
	  • Security incidents or violations reported, e.g. virus infection incidents (from incidents logbook).
	  • Request to visit or access and surfing of unauthorized online services and websites.
	  • Use of weak passwords.
	  • Not installed or disabled anti-virus software.
	  • Sending of sensitive information via email.
	  • Personal data disclosure or breach.
	• Tool-based attack, e.g. to crack a password and measure the strength of passwords created before and after the awareness program.
	• Silent observation of compliance (in an organization, preferably, after work hours)

Attitude	**Intrusive method** • Questionnaire-based survey (quantitative or qualitative; open-ended questions), semi-structured interview, and group discussion to know wishes, concerns, problems, values, beliefs, norms, and willingness of cybersecurity. **Non-intrusive method** • System data (interest in an awareness program), e.g. count of information security intranet page accesses, or visits to a webpage where awareness information is uploaded. • Security related helpdesk calls, i.e. count of calls to helpdesk that run counter to the purpose of awareness-raising. • Silent observation of security-related activities.
Knowledge and competence	**Intrusive method** • Standardized survey questionnaire to measure knowledge and competence. • (Pre- and post-) tests using vocabulary and scenario type questions, e.g. phishing screenshots to identify.
Interest	**Interest by audience** • Survey (quantitative or qualitative) and other qualitative approaches, e.g. interviews and group discussion. • Percentage of attendees (i.e. attendance) with respect to the expected number of attendees (if mandatory in the organization, most employees are forced to attend, and may not represent the real interest; voluntary participation shows the real interest. • Silent observation of participants during the session, e.g. yawning, side talking, frequency of short breaks taken. **Interest by organizer** • Motivation demonstrated (observation) by those playing key roles in managing/coordinating cybersecurity program. **Interest by management** • Moral support and commitment by management (observation and interview) for an awareness program. • Fund and resources allocated for an awareness program, e.g. to support distribution (i.e. use of dissemination channels) and posting of security awareness items.
Reachability	Accessibility of awareness materials. • Survey to know who received the awareness information. • Percentage of people who attended an awareness session. • Count of people that received a leaflet. • Number of attendees visiting the e-learning program in e-learning. • Count of email recipients. • Count of people logged into iNotice. • Visit the website (but can have repetitive visits from a small group of people.

Table 4: Continued

Measured factor	Measurement method
Touchability	Self-motivated actions. • Feedback forms, survey (anonymous), interviews, and focus group discussions. • Awareness or security day communication (face-to-face feedback). • Attendance when it is not mandatory. • Number of attendees who registered and completed the program in e-learning. • Hit counts to the link for more information in the email. • Hit counts to the link for more information in iNotice. • Posters downloaded from the link provided. • Simulated attacks and response observation. • Independent observations (e.g. awareness of clean desk policy, observation performed outside working hours). • Comparisons of pre- and post-session tests or survey results. • Audience satisfaction (e.g. attendees temporarily leaving the room, constantly chatting with colleagues, or sketching on their notes are not encouraging reactions to be discouraged. • Visits to shared information. • Independent observations (behavior. • Audit and risk department reports (count of security issues related to employees). • Security incidents reported.
Value-added	**Non-financial benefit** • Contribution is realized based on recognition of security contributions, e.g. count and reputation of awards and contests won. • Percentage of awareness processes incorporated in the organization's processes. **Financial benefit** • Financial cost calculation of organizing an awareness program.
Usability	**Relevant topics covered** • Percentage of relevant security topics covered (with respect to expected topics to cover). • Survey, interview, and group discussion to realize the covered topics were suitable for the audience. **Delivery assessment** • Post-awareness survey to know the learning method was preferred by the audience. • Pass and fail rates, frequency of awareness program, and count of attendees. **Usage of knowledge in practice** • Survey using a questionnaire to know the usage of knowledge in practice. **User confidence and satisfaction** • Survey using a closed questionnaire. • Satisfaction measured using a qualitative approach: interviews, group conversations, and observation. • Post-awareness questionnaire to realize confidence, satisfaction, and preferences. • Users' exposure to awareness materials is increasing. • Users with significant security responsibilities being appropriately trained is increasing. • Coverage and identified needs are shrinking. **The usefulness of the awareness program** • Survey using a closed questionnaire. • Semi-structured interview to realize the percentage of the audience that found the organization of the event satisfactory (i.e. suitability and importance of the issues discussed, program organization, and program duration).
Overall feedback	**Feedback strategies** • Post-event survey that can be qualitative or quantitative (preferably anonymous. • Feedback forms (preferably anonymous). • Focus group discussion. • Selective/informal interview. • Informal break room conversation.

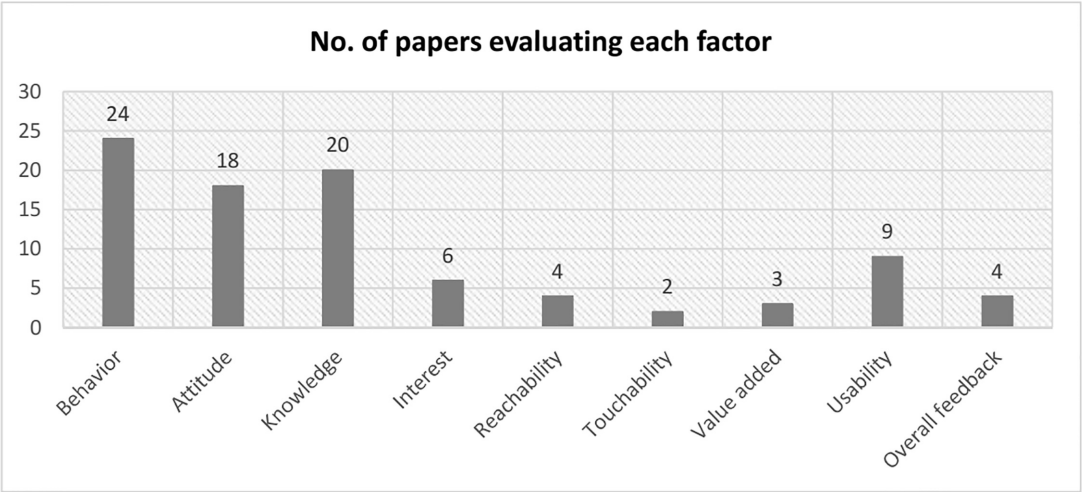

Figure 2: Factors measured by the reviewed papers

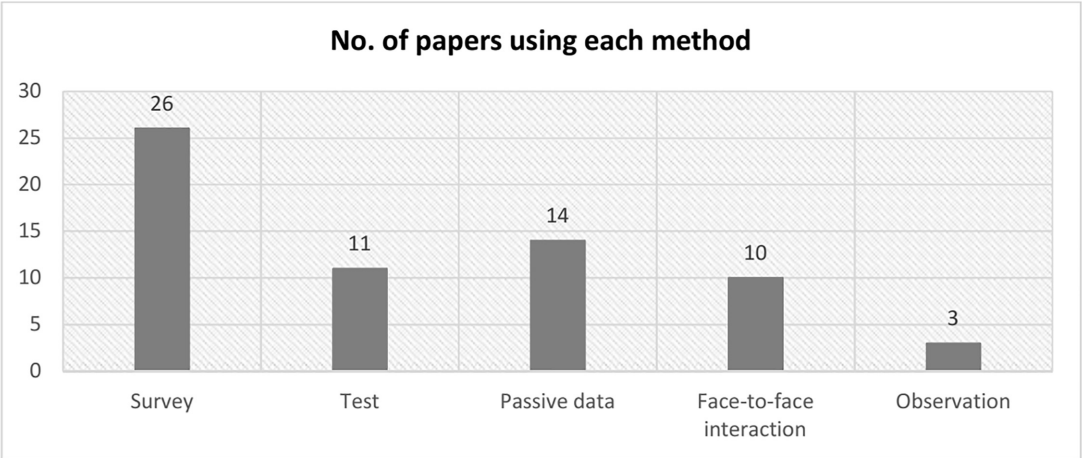

Figure 3: Methods used for the evaluation CSA

the participants are aware of the things happening to and around them, which could lead them to be alert or possess a high cognitive load. Such participants (alert or with a high cognitive load) may not think and act in the ways they would do in a natural environment. This could negatively affect the accuracy of the outcomes. The evaluation is generally conducted in a laboratory environment using well-established evaluation techniques, such as questionnaires, interviews, tests, observations in a laboratory setting, and focus group discussions. Its main advantage is a short evaluation duration (a session may last from a few hours to a day) within which different aspects of cybersecurity can be covered. An intrusive evaluation could be suitable for assessing/measuring security knowledge and attitude, where the familiarity with facts and information, and individual viewpoints are measured, respectively.

Whereas in non-intrusive evaluation, the participant's normal behavior is not consciously disrupted by the evaluation processes. The evaluation processes are performed in a natural environment with a lower cognitive load in the participants, which may result in more accurate outcomes. However, the main disadvantages of this evaluation are a longer evaluation period (can continue for a few days depending on the number of participants and the aspects to be assessed), only a few existing evaluation techniques are applicable (e.g. analysis of system and log data, disguised observations, simulated and other attacks), and the likelihood of bias introduction is high that may skew the outcomes (e.g. in observation the evaluator has to interpret the participant's activities). A non-intrusive evaluation could be more appropriate to assess security behavior. Moreover, it should be preferred to assess/measure other factors wherever possible in organizations since it will save the participant's (or employee's) effort and time. However, sometimes getting an in-depth understanding of why the participants behaved in a certain way may require using a follow-up intrusive method.

The widely used methods for evaluation purposes are categorized broadly in Fig. 3 and explained next. Although some of the categories overlap with each other based on their true definitions, we have distinguished them in the description to improve their clarity and understanding.

Survey

A questionnaire survey has been found to be the most popular method used. It has been implemented mainly to determine the im-

pact of a CSA program on the security knowledge, attitude, and behavior of the participants. In order to do so, pre- (before the awareness program) and post- (after the awareness program) surveys have been utilized. In addition to these factors, post-survey has also been carried out for many other purposes, such as

- To determine the suitability and usefulness of the covered topics.
- To realize the importance of the knowledge gained in practice.
- To understand the interest and willingness to participation.
- To realize the confidence, satisfaction, and preference of the audience.
- To determine preference for learning methods.
- To get overall feedback (or suggestion, opinion).

This popularity of a questionnaire survey could be because (i) it allows a large population to be assessed with relative ease, (ii) it is easy to integrate different aspects of cybersecurity for evaluation in a survey, and (iii) it is possible to reduce time, effort, and cost for conducting a survey by using techniques like an online survey (it is economical to disseminate questionnaires, and data gathering is automated), survey sampling (it reduces sample size when the target size is large), and quantitative questions (it is relatively easy, and fast to analyze quantitative questions).

A survey can use closed-ended questions, open-ended questions, or a mixture of both question types. When close-ended questions are used, the results can be quantified; however, open-ended questions are presumably more suitable to get deeper insights into attitudes and behavior changes [45] if the respondents are adequately literate and interested to answer. On the flip side, responses to open-ended questions may be difficult to interpret and analyze, limiting their usefulness.

Finally, a survey captures active data from willing participants, who may not remember correctly all the things they do. Such memory lapse can introduce incorrect data that ultimately could restrict the ability to use the information for actionable results. This can be mitigated to an extent by using closed-ended questions whenever possible, where all possible options are provided, making it easier for the participants to remember and recall their preferences. Moreover, making the survey anonymous can encourage the participants to provide real and honest feedback. These all are important since the survey result is of value as long as the participants do not lie.

Passive data

Analyzing passive data (also referred to as indirect observation) has been found to be the next popular method used mainly to evaluate the behaviors of the participants and their interest in CSA. When gathering data through direct observation becomes very expensive and time-consuming, then indirect observation becomes more relevant. The reviewed studies have collected passive data from multiple sources, such as the audit department, the risk department, other external and internal auditors, and the helpdesk in a natural environment (i.e. participants remain unaware of the data collection for a research purpose). This data is not subjective (i.e. human independence, so a separate time of the audience for the data collection is not required), and easy as well as economical to obtain. These may be some reasons why many studies utilized them to evaluate cybersecurity behaviors (both risky behaviors and best security or compliance behaviors). Some types of passive data used to measure cybersecurity behavior are

- Anti-virus and firewall logs.
- Visits or requests to visit unauthorized services and websites.
- Number of security incidents or violations reported.
- Use of weak passwords.
- Sending of sensitive information *via* email.
- Count of calls to the helpdesk.
- Visits or traffic to the location where awareness information is available (e.g. security intranet page, or location where awareness information is uploaded).
- Click through rate of malicious links.
- Count of information security intranet page access, or visits to a webpage where awareness information has been uploaded.
- Installation/non-installation of security protection.
- Coverage and identified needs of CSA are shrinking.
- Frequency of awareness programs needed in the organization.
- Increase in reporting of potential cyber incidents by cyber aware people.

Similarly, passive data that has been utilized to know whether the CSA information reached the target audience or not are

- Count of people that received a leaflet.
- Number of attendees visiting the e-learning program.
- Count of email recipients.
- Count of people logged into iNotice.
- Visits the website (but there is a risk that a small group of people may repeatedly visit the website).
- Percentage of people who attended an awareness session.

The audience's interest in a CSA program (or whether cybersecurity information touched the audience or not) is also determined by utilizing the following passive data:

- Number of attendees registering, and completing the e-learning program.
- Hit counts to the link for more information in the email/iNotice.
- Poster downloaded from the link.
- Activities like attendees temporarily leaving the room, constantly chatting with colleagues, or sketching on their notes are not encouraging reactions.

As a matter of fact, utilizing such data for evaluation purposes in cybersecurity provides a more realistic outlook for the situation. This data is a part of everyday activities, so participants do not need a separate notification that could make them aware and alert, thus influencing their activities and data.

However, analyzing this data could answer only what has happened and not why has it happened, since the evaluator gathers information without the direct involvement of the people studied. For example, it will answer if the participants continue to use a weak password even after participating in a related awareness program, but not provide an answer to why they continue doing so. Indeed, it is important to know whether an awareness program has brought positive changes in the participant's behavior or not, but at the same time, it is equally necessary to know why certain behaviors have not changed. This information will help to improve the awareness program in a future iteration. In addition, access to someone's digital footprints will require a precautionary approach, e.g. implementing privacy-enhancing technologies, incorporating regulatory controls, and receiving permission from the authority.

Test

Tests in two forms have been utilized for evaluation purposes, which are (i) a question-based test, and (ii) an attack-based test. Such tests are performed before and after a CSA program and their results are compared to know the effectiveness of the CSA program. These tests are conducted mainly to evaluate cybersecurity knowl-

edge, and in the case of a simulated attack, are also used to evaluate behavior.

In a question-based test (e.g. quiz, or game) using standardized questions [46] comprising vocabulary and scenario type questions [36] can help to ask the right and relevant questions. Then, in an attack-based test, using a secret simulated attack, e.g. sending a phishing email to the audience and observing their responses like a count of people who revealed sensitive information, who downloaded or opened the attachment, who identified and reported the phishing attempt to the concerned authority, and who reported about fallen for a phishing attack after realizing it, can provide more realistic results. Similarly, other attack types, such as checking password strength using tools and techniques after an awareness program on creating strong passwords, can also be a test to evaluate the effectiveness of the awareness program.

As with the passive data approach described above, this simulated attack also provides a more realistic view of the situation. But it involves more work (like developing attacks in as natural form as possible, taking care of legal and ethical aspects, and others) and can be expensive to conduct. Further, for various aspects of cybersecurity, other forms of evaluation could be more appropriate than using simulated attacks, e.g. the observation method would be suitable to determine whether an individual leaves his/her digital devices unattended, or passive data would be suitable to know whether an individual routinely updates the anti-virus software in his/her digital devices. More importantly, exposing people to simulated attacks can have several unintended consequences (negatively impact the staff trust, and security and error culture of an organization) and could also violate various national or data protection laws, or local agreements [86]. For example, a phishing attack attempts to persuade victims to reveal sensitive information, download and open a malicious attachment, circumvent security in digital devices, transfer money, and so will be a simulated phishing attack that may not be compatible with different laws and agreements. Therefore, while conducting such attacks, it is mandatory to ensure that no laws and ethics are contravened.

Face-to-face interaction

Face-to-face interaction using techniques like semi-structured interviews, informal break room conversations, and focus group discussions to get audience and management feedback on a CSA program has been found to be the next popular method. It can be either targeted or generalized (e.g. suggestions, opinions, wishes, concerns, problems, and values). Such face-to-face interaction conducted in a laboratory setting can also be utilized to an extent to realize the audience's cybersecurity knowledge, attitude, and behavior.

One of the main advantages of face-to-face interaction is that it captures both verbal and non-verbal (e.g. nuances of the voice, facial expressions) cues. In addition, it supports immediate feedback and without any delay clears up confusion and misunderstanding in the message conveyed if any exists. However, in a face-to-face interview, the cost can be a major disadvantage since it requires an interviewer (i.e. personnel cost) and is very time-consuming to conduct. Its cost can be reduced by using informal break room conversations and focus group discussions, although they may not provide a comprehensive understanding of the problem. Also, face-to-face interaction does not provide anonymity, which can be a concern for some respondents.

Observation

Both disguised (i.e. evaluator's presence is concealed from participants) and non-disguised (i.e. evaluator's presence is known to participants) observations have been found to be utilized mainly to evaluate cybersecurity behavior. Observation can be both direct (involves looking at the actual behaviors) and indirect (involves looking at a result of behaviors). This subsection, by observation, refers only to the direct one. The indirect observation has been included as passive data in subsection Passive data.

Although the non-disguised method alleviates ethical concerns that may arise due to watching someone covertly, it suffers the *Hawthorne effect* [87], i.e. participants act differently when they are being watched, and could not provide the actual behavioral changes resulted due to an awareness program. In addition, there is always a risk of distracting and disturbing the participants from their normal activities. At the same time, the non-disguised method can be replaced with a survey that can be equally effective. In that case, the disguised method conducted in the natural environment is a more preferable method for assessing the behaviors of the participants. But this also has a downside, which is due to the absence of interaction between evaluator and participants, there is a high chance that the evaluator may introduce errors and bias in the analysis of behavioral events. Moreover, in the case of an organization where the participants are within a specified perimeter, conducting a disguised observation could be easy, but doing the same may not be feasible for an awareness program that targets the general public.

There often exists a disconnect between what people self-report they do and what they actually do. So, to study a change in the participant's behaviors after participating in a CSA program, observation can be a very effective method. A more focused or structured observational study (where the evaluator uses checklists or targets specific behaviors) can be a more dependable method. This requires the evaluator to know what to observe (*event sampling*), when to conduct the observation (*time sampling*), and how to *document* the observations. The evaluator is often suggested to record the events for discussion and analysis at a later stage.

But a major limitation of observation is that it is generally conducted in-depth over a prolonged period, with data that are often subjective and difficult to quantify, thus the sample size is usually kept at a minimum. Moreover, it requires skilled observers and analysts, otherwise they may introduce errors and biases in their analysis. Not to mention, it does not provide anonymity to the participants.

Metrics Development

But prior to metrics development, it is important to realize what constitutes good metrics. Some criteria of good metrics, which we believe are relevant for our proposed metrics [13] are shown in Fig. 4.

Evaluation can be diagnostic (i.e. a pre-assessment conducted to know an audience's existing awareness level on the topic), formative (i.e. an assessment conducted during the program development and implementation to realize the needs and processes required to achieve the goal), and summative (i.e. a post-assessment conducted to assess the outcome of the program and determine broader and long-term changes occurred due as a result of the program). The diagnostic assessment followed by the summative assessment is mainly related to the outcome and impact of the program or the declaration of the success or failure of the program, but the formative assessment helps learn where to best put the limited resources available for CSA. For a complete evaluation of a CSA program, all three assessments are equally necessary.

For the evaluation purpose, it is imperative to have a *clear* goal [5] and *measurable* objectives [88] from a CSA program. More essentially, both the goal and objectives must be realistic or achievable.

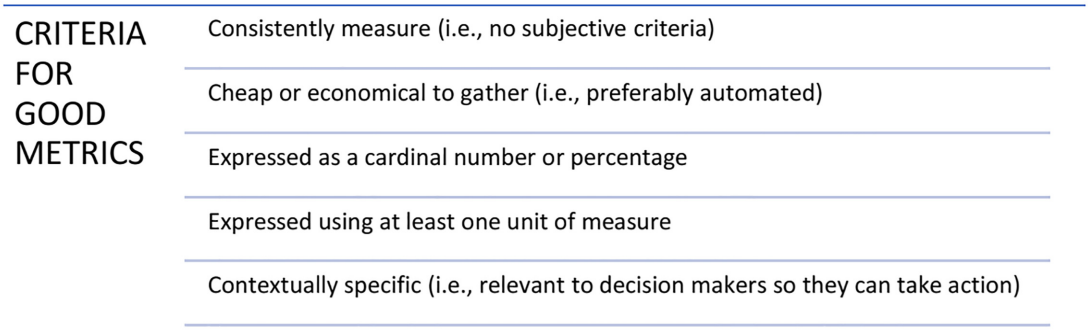

Figure 4: Criteria for good metrics [13]

The goal and objective should serve to uphold the reason for creating an awareness program, i.e. what the program wants to achieve. They can be unique to each target group. The *measurable* objective can be tracked with the help of numbers and units, which is crucial for continuously monitoring and analyzing the success. Against this objective, the effectiveness of the program is evaluated, and accordingly, the program is revised and updated.

In general, a CSA program is expected to communicate cybersecurity knowledge (i.e. recommended guidelines and security best practices) to the target audience, broaden the cybersecurity knowledge of the target audience (i.e. familiarity with guidelines and security best practices), bring positive changes in attitude (i.e. motivate to adopt recommended guidelines and practices) and behavior (i.e. create a strong culture of security) in the target audience, gain and keep the audience and management/sponsor trust and satisfaction; and ultimately minimize the number and extent of security breaches [5]. But these expectations are difficult to quantify. Some examples of clear goals and their respective measurable objectives are as follows:

- ***Goal***: achieve compliance with required regulations and directives; ***objective***: compliance with GDPR, e-Privacy Regulation, NIS Directive, and so on.
- ***Goal***: identify and manage human risks to an acceptable level; ***objectives***: reduce accidental data loss incidents by 70%.
- ***Goal***: raise awareness of security best practices; ***objectives***: use of password security, practice social media safety, practice malware protection, practice mobile security, awareness of phishing, and so on.

In addition, the evaluation process must be cost-effective (or inexpensive) to conduct and its results are useful for decision-making. Cost can be reduced by limiting to only variables that need to be measured and doing this in a more planned and structured way in terms of schedule and clarity in questions intended to be answered. Then, usefulness can be improved by understanding and taking into account the priorities and concerns of different stakeholders who will use the evaluation findings for decision-making.

The ELINET [89] recommends four indicators and their measurement methods that are important for the evaluation of awareness activities. Based on this recommendation, the aforementioned criteria for good metrics, and evaluation methods utilized by the reviewed studies, we propose the metrics as shown in Fig. 5 and explained in Table 5 for the evaluation of a CSA program. We believe that all these four indicators are important to be evaluated in order to know the effectiveness and success of a CSA program. It is possible that an organization may not be in a situation to afford the measurement of every factor. In a situation like this, it is suggested that the organization measure selective factors most relevant to it from each indicator rather than measuring all factors from a certain indicator while abandoning other indicators. The target audience will impact how each indicator can be measured. For example, it may be easy and economical to obtain system and log data if the target audience is the organizational staff (they are in a controlled environment), but such data may not always exist if the target audience is customers (they are in an uncontrolled environment). Moreover, while suggesting measurement/assessment methods, we have tried to ensure that they adhere to the criteria for good metrics. For example, we have emphasized a quantitative method, i.e. non-subjective as well as quantifiable, and so makes sense to the sponsor/management. Besides, we have provided multiple alternatives to measure each indicator type so that the cost-effective option can be selected.

Furthermore, an evaluation should not be limited to what factors to measure and how to measure them, but should also cover whom they have been measured for. This will help in the complete evaluation (i.e. from the perspective of all important stakeholders like CSA professionals, management/sponsors, and an audience group) and at the same time provide an idea of who will participate in the evaluation process. Outcomes from the evaluation of impact factors and accessibility factors are more connected to the CSA professionals, who are responsible for updating and adjusting the CSA program for future iterations. Whereas evaluation results of sustainability indicators and monitoring indicators are helpful for the management or sponsor in deciding whether to continue investing in the existing awareness program or have to look for an alternative.

Other important aspects are a proper visualization of the evaluation results and automation of the measurement processes [18]. For the visualization of evaluation results, a scorecard can be a potential option [5, 18]. Likewise, several tools are available that can be used to automate data collection processes using surveys, tests, and passive data; however, to use them could demand a certain set of technical skills.

Different from the works discussed in the section Related Works, our proposed metrics have given equal importance to the evaluation of the sustainability indicators. Sustainability can be expressed in terms of the program outcome's ability to exist constantly by influencing organizational policies, arrangements, and regulatory framework. It can also be expressed in terms of the program's ability to exist constantly in the organization by becoming a part of the organizational policies and receiving abundant funds. CSA is a continuous process, and the evaluation process is similarly iterative. Without the evaluation of the sustainability indicator, the continuity of the process itself can become questionable. More importantly, the evaluation should seek input from all of those involved and affected by a CSA program. This is possible only by ensuring that diverse viewpoints from different stakeholders are considered so that the re-

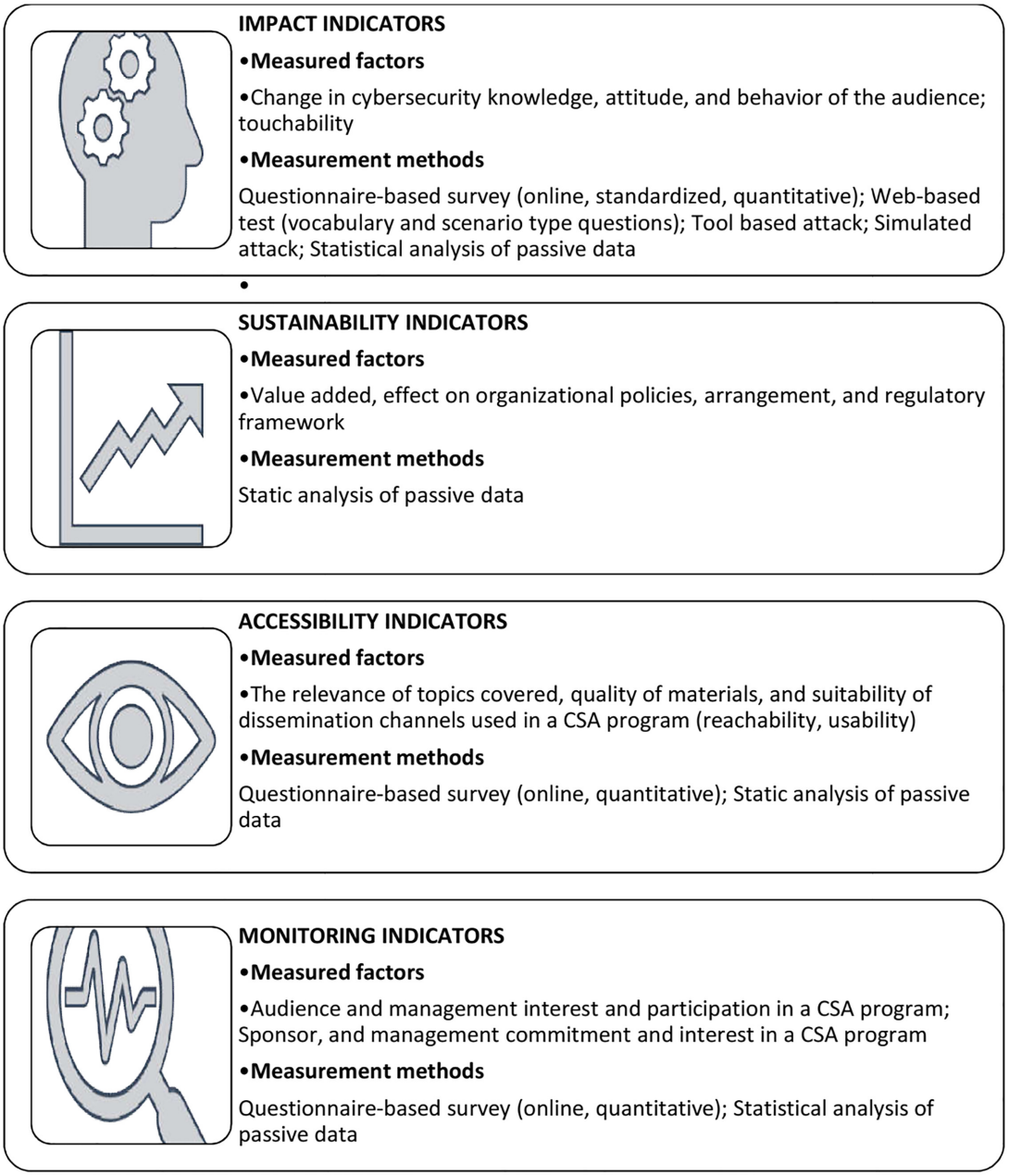

Figure 5: Metrics for the evaluation of a CSA program

sults are as complete and unbiased as possible. Ironically, none of the work discussed in the section Related Works has clearly considered this aspect and has based its evaluation completely on the audience's viewpoint. Indeed, the audience assessment and feedback are important, but to yield a more balanced and holistic picture of the awareness program, it is necessary to measure how well the expectations of each stakeholder are met. This can provide critical insight and guidance to adjust the awareness program for future iterations.

The proposed metrics can guide the evaluation process of a CSA program; however, they do not answer what score is an acceptable level of awareness [21, 26]. This is an important question, but is contextual and will vary depending on the target topic and audience type. For example, if the target audience is healthcare or banking staff, the only acceptable score will presumably be the maximum. Therefore, it is necessary to set a benchmark expectation from the CSA program [4, 5]. Further, for any evaluation, there needs to be clarity about what will be considered a quality and ethical evaluation [90]. This could vary depending on organizational policies, laws, and regulations. Some organizations could have in place particular evaluation standards and/or ethical guidelines to guide the evaluation. The most important thing is to approach it methodically and attentively to avoid any unforeseen repercussions, unnecessary interruptions, or meaningless outcomes.

Table 5: Metrics for the evaluation of CSA

Indicator	Measured factor	Measurement/assessment method

Impact indicators
Measure and assess the learning (i.e. knowledge and skills gained by the audience as a result of the awareness), and the impact on the audience's performance and attitude toward cybersecurity.
	Impact of awareness on:
	• Cybersecurity knowledge and competence.
	• Attitude to cybersecurity.
	• Cybersecurity behavior.
	It also comprises touchability.
		• (Pre- and post, quantitative) web-based test (vocabulary- and scenario-type questions) to determine if the audience knows more about the issues covered by the awareness program or not.
		• (Pre- and post, online, standardized, and quantitative) questionnaire-based survey to determine if the audience knows more about the issues covered by the awareness program or not, and understands the sense of urgency of fighting and preventing the issue or not.
		• (Pre and post) statistical analysis of passive data to know if there is a decline in security incidents and violations, for example:
		  ○ Data from audits and risk departments.
		  ○ Count and severity of security incidents occurred due to staff behaviors.
		  ○ Other best behavior data that can be automatically collected (e.g. anti-virus and firewall log data, and helpdesk data).
		• (Pre and post) simulated and tool-based attack to determine if the audience understands the sense of urgency of fighting and preventing the issue or not.

Sustainability indicators
Measure the direct and indirect values added to the organizations as a result of implementing CSA. These indicators are critical for the management or sponsors in their decision-making on whether to invest in the program or not, and this is necessary for the continuity of the program.
	Impact of awareness in the change of:
	• Organizational policies.
	• Regulatory framework.
	• Organizational arrangement.
	Change in top management and sponsor support and commitment for the awareness program
		• Valued-added by the awareness program evaluation based on, for example:
		  ○ Recognition of security contributions, e.g. count and reputation of awards and contests won due to the awareness program.
		  ○ Percentage of awareness processes incorporated in the organization's policies, processes, and arrangement
		• Change in funding and resources allocated for the awareness program to realize the management/sponsor interest in the awareness program.
		• Cost-benefit analysis of the program (i.e. ROI).

Accessibility indicators
Measure the quality of resources and delivery channels used in the awareness program.
	Quality of awareness resources. Effectiveness of awareness resources.
	For example, whether the content was relevant and easy to follow or not, what were the strengths and weaknesses of the program, and whether the delivery methods were able to accommodate the audience's pace and learning style or not. It comprises of usability and reachability.
		• Survey to evaluate (using closed questions/quantitative, such as Likert scale).
		  ○ Relevancy of topics.
		  ○ Content quality.
		  ○ Delivery assessment.
		• Percentage of security topics covered with respect to expected topics to be covered to know if all relevant or demanded topics are covered or not.

Table 5: Continued

Indicator	Measured factor	Measurement/assessment method
		• System and log data analysis (e.g., attendance, website visit, email recipient, etc.) to determine if the target group has access to the awareness resources or not.
Monitoring indicators Measure how the audiences, sponsor, and senior management have perceived or reacted to the awareness program.	Interest, support, commitment, and participation of different stakeholders in the program.	Interest and active participation evaluated using: • System and log data analysis (e.g. attendance when it is not mandatory, number of attendees who registered and completed the e-learning program with respect to those who visited, hit counts to the link for more information, and so on). • Post-event survey (using closed questions/quantitative, such as Likert scale; preferably anonymous) to receive overall feedback on the awareness program. • Availability of resources for the program.

Conclusions

The evaluation of a CSA program is an important activity in the post-implementation phase. Evaluation is necessary to know how effective and successful the program was. Moreover, it provides information on which aspects of the program require improvement and also information used by senior management/sponsor in deciding whether to invest further in the program.

In spite of all the benefits of evaluation, there does not exist a consensus on what to measure and how to measure while evaluating a CSA program. This may be because different target groups have varying needs and environments determining the content of their CSA programs; so generalized evaluation metrics cannot capture the rationale behind an evaluation strategy. Ironically, this lack of evaluation metrics for CSA has caused more harm than good: e.g. many organizations and individuals either abandon the evaluation process or limit their evaluation to some weak or irrelevant factors and indicators. Therefore, in this paper, we have designed and proposed evaluation metrics for CSA that we believe are widely applicable.

In order to do so, we performed a systematic literature review of 32 past studies that have evaluated or proposed methods to evaluate a CSA program. We gathered the relevant papers after multiple rounds of screening. A review of the gathered papers followed this, mainly to extract information on what factors past studies measured and how they measured them to evaluate or assess the effectiveness and success of a CSA program. Analysis of the collected data revealed that factors measured by the past studies can be classified into behavior, attitude, knowledge, interest, reachability, touchability, value-added, usability, and overall feedback. Among all the factors measured, behavior, attitude, and knowledge are the most popular factors. Similarly, methods used to measure these factors can be categorized into a survey, test, passive data, face-to face-interaction, and observation, where survey and passive data are found to be the most popular.

Using the obtained findings, criteria for good metrics, and the ELINET's four indicators (i.e. impact, sustainability, accessibility, and monitoring), we have designed and proposed new metrics for the evaluation of a CSA program. Our proposition provides factors to be measured and their respective measurement methods in order to realize each of the indicators.

The impact indicator is realized by measuring positive changes in cybersecurity knowledge, attitude, and behavior due to the CSA program using methods like online surveys, tests (web-based, tool-based, and simulated), and statistical analysis of relevant passive data.

Similarly, the sustainability indicator is realized by measuring the changes in organizational policies, regulatory framework, and organizational arrangement due to CSA. Moreover, it is also realized by measuring the change in senior management and sponsor support, and commitment to CSA. The sustainability indicator is measured using statistical analysis of relevant passive data like the percentage of awareness outcomes integrated into the organizational process, policy, and arrangements; cost-benefit analysis; and changes in funds and resources allocated for the program.

Next, the accessibility indicator is realized by measuring the relevancy of topics, quality of materials, and appropriateness of delivery channels using methods like surveys, the percentage of relevant topics covered, and statistical analysis of relevant passive data like audience interest in the awareness program.

Finally, the monitoring indicator is realized by measuring stakeholders' interest and participation in the awareness program using passive data analysis and post-program surveys.

We believe our proposition is inclusive of all directly affected stakeholders, i.e. management, CSA professionals, and target audiences. More importantly, the proposed metrics have considered various important aspects, such as criteria for good metrics, different stakeholder needs, and the sustainability of the program in order to make the evaluation process inclusive, complete, and unbiased as far as possible.

Acknowledgments

The authors would like to thank David Goodman (Trust in Digital Life, Belgium) and Pasquale Annicchino (Archimede Solutions SARL, Switzerland) for reviewing the deliverable report submitted to the CyberSec4Europe.

Funding

This work has financially been supported by the CyberSec4Europe project (grant agreement no. 830929). This paper is an extended and revised version of the deliverable report [32] that was submitted to CyberSec4Europe WP9: dissemination, outreach, spreading of competence, and raising awareness.

References

1. Williams S. More than half of personal data breaches caused by human error. IT Brief. 2019.
2. IBM Global Technology Services, *IBM Security Services 2014 Cyber Security Intelligence Index Analysis of cyber attack and incident data from IBM's worldwide security operations*. Somers: IBM Corporation. 2014.
3. Kaspersky. The human factor in IT security: How employees are making businesses vulnerable from within. Kaspersky Daily. 2018.
4. Wilson M, Hash J. Building an information technology security awareness and training program. https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-50.pdf. (20 September 2020, date last accessed).
5. ENISA. The new users' guide: How to raise information security awareness. https://www.enisa.europa.eu/publications/archive/copy_of_new-users-guide/at_download/fullReport. 2010. (20 September 2020, date last accessed).
6. Hänsch N, Benenson Z. Specifying IT security awareness. In: *Proceedings of the Twenty-fifth International Workshop on Database and Expert Systems Applications*, Munich, 2014.
7. Bada M, Sasse AM, Nurse JR. Cyber security awareness campaigns: why do they fail to change behaviour?. In: *Proceedings of the International Conference on Cyber Security for Sustainable Society*, Coventry, 2015.
8. Katsikas S. Health care management and information system security: awareness, training or education?. *Int J Med Inf* 2000;**60**:129–35.
9. McCrohan KF, Engel K, Harvey JW. Influence of awareness and training on cyber security. *J Internet Commer* 2010;**9**:23–41.
10. Furnell S, Vasileiou I. Security education and awareness: just let them burn?. *Netw Secur* 2017;**2017**:5–9.
11. ENISA. Information security awareness initiatives: current practice and the measurement of success. https://ifap.ru/library/book206.pdf. (20 September 2020, date last accessed).
12. Rohlich N, Haas P, Edwards F. Exploring the effectiveness of transit security awareness campaigns in the San Francisco Bay area. https://transweb.sjsu.edu/research/Exploring-Effectiveness-Transit-Security-Awareness-Campaigns-San-Francisco-Bay-Area. (20 September 2020, date last accessed).
13. Spitzner L. Security awareness metrics. https://www.sans.org/security-awareness-training/blog/security-awareness-metrics. (20 September 2020, date last accessed).
14. Timmermans B, Cleeremans A. How can we measure awareness? An overview of current methods. In: Overgaard M (ed.), *Behavioural Methods in Consciousness Research,* Oxford: Oxford University Press, 2015, 21–46.
15. Fogg B. A behavior model for persuasive design. In: *Proceedings of the Fourth International Conference on Persuasive Technology*, Claremont, CA, p. 26–9. 2009.
16. Richardson R. CSI computer crime & security survey. http://www.sis.pitt.edu/jjoshi/courses/IS2150/Fall11/CSIsurvey2008.pdf. (20 September 2020, date last accessed).
17. Monaha D. *Security Awareness Training: It's Not Just For Compliance*. Boulder: Enterprise Management Associates (EMA), 2014.
18. Fertig T, Schütz AE, Weber K. Current issues of metrics for information security awareness. In: *Proceedings of the Twenty-Eighth European Conference on Information Systems*, Virtual conference, Marrakech, p. 15–7. 2020.
19. Spitzner L, deBeaubien D, Ideboen A. Security awareness report. Bethesda, MD: SANS Institute, 2019.
20. Dixson DD, Worrell FC. Formatibe and summative assessment in the classroom. *Theory Into Practice* 2016;**55**:153–9.
21. Manifavas C, Fysarakis K, Rantos K. *et al.* DSAPE: dynamic security awareness program evaluation. In: *Proceedings of the Sixteenth International Conference on Human-Computer Interaction*, Crete, p. 258–69. 2014.
22. Bada M, Nurse JRC. Developing cybersecurity education and awareness programmers for small and medium-sized enterprises (SMEs). *Inf Comput Secur* 2019;**27**:393–410.
23. Gattiker UE. Can an early warning system for home users and SMEs make a difference? A field study. In: *Proceedings of the International Workshop on Critical Information Infrastructures Security*, Samos Island, 2006.
24. Shaw R, Chen CC, Harris AL. *et al.* The impact of information richness on information security awareness training effectiveness. *Comput Edu* 2009;**52**:92–100.
25. Bitton R, Boymgold K, Puzis R. *et al.* Evaluating the information security awareness of smartphone users. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI'20)*, Honolulu, HI, p. 25–30. 2020.
26. Kruger H, Kearney W. A prototype for assessing information security awareness. *Comput Secur* 2006;**25**:289–96.
27. Webster J, Watson RT. Analyzing the past to prepare for the future: writing a literature review. *MIS Quart* 2002;**26**:xiii–xxiii.
28. Google Scholar. https://scholar.google.com/intl/en/scholar/about.html. (7 July 2021, date last accessed).
29. Microsoft Academic. https://academic.microsoft.com/faq. (7 July 2021, date last accessed).
30. Kitchenham B *Procedures for Performing Systematic Reviews*. Keele: Software Engineering Group, Department of Computer Science, Keele University, 2004.
31. Paez A. Gray literature: an important resource in systematic reviews. *J Evid Based Med* 2017;**10**:233–40.
32. Chaudhary S, Gkioulos V. D9.13 Awareness effectiveness study. 2021. (16 February 2021, date last accessed).
33. Chaudhary S, Gkioulos V, Goodman D. D9.11: SME cybersecurity awareness program 2. https://cybersec4europe.eu/wp-content/uploads/2021/05/D9.11-SME-cybersecurity-awareness-program-2-FINAL-submitted-1.pdf. (14 July 2021, date last accessed).
34. Dodge RC, Ferguson AJ. Using phishing for user email security awareness. In: Security and Privacy in Dynamic Environments. *Proceedings of the IFIP TC-11 Twenty-First International Information Security Conference (SEC 2006)*, Karlstad, p. 22–4. 2006.
35. Kruger HA, Drevin L, Steyn T. A framework for evaluating ICT security awareness. In: *Proceedings of the ISSA 2006 from Insight to Foresight Conference, July 5-7*. Sandton. 2006.
36. Kruger H, Drevin L, Steyn T. A vocabulary test to assess information security awareness. *Inf Manag Comput Secur* 2010;**18**:316–27.
37. Albrechtsen E, Hovden J. Improving information security awareness and behavior through dialogue, participation, and collective reflection: an intervention study. *Comput Secur* 2010;**29**:432–45.
38. Khan B, Alghathbar KS, Nabi SI. *et al.* Effectiveness of information security awareness methods based on psychological theories. *Afr J Bus Manag* 2011;**5**:10862–8.
39. Wolf M, Haworth DA, Pietron L. Measuring an information security awareness program. *Rev Bus Inf Syst* 2011;**15**:9–22.
40. Ahlan AR, Lubis M. Information security awareness in university: maintaining learnability, performance, and adaptability through roles of responsibility. In: *Proceedings of the Seventh International Conference on Information Assurance and Security (IAS)*, Melaka, p. 5–8, 2011.
41. Tsohou A, Karyda M, Kokolakis S. *et al.* Analyzing trajectories of information security awareness. *Inf Technol People* 2012;**25**:327–52.
42. Bauer S, Bernroider EW, Chudzikowski K. End user information security awareness programs for improving information security in banking organizations: preliminary results from an exploratory study. In: *Proceedings of the AIS SIGSEC Workshop on Information Security & Privacy (WISP2013)*, Milano. 2013.
43. Gundu T, Flowerday S. Ignorance to awareness: towards an information security awareness process. *South Afr Inst Elect Eng* 2013;**104**:69–79.
44. Velki T, Solic K, Ocevcic H. Development of user's information security awareness questionnaire (UISAQ). In: *Proceedings of the International Convention MIPRO*, Opatija, p. 26–30. 2014.
45. Prah ANW, Otchere AA, Opan KE. The perceived effectiveness of information security awareness. *Inf Knowl Manag* 2016;**6**:62–73.

46. Scholl MC, Leiner B, Fuhrmann F. Blind spot: do you know the effectiveness of your information security awareness raising program?. *Syst Cybernet Inf* 2017;**15**:58–62.
47. Carella A, Kotsoev M, Truta TM. Impact of security awareness training on phishing click-through rates. In: *Proceedings of the IEEE International Conference on Big Data*, Boston, MA, p. 11–4. 2017.
48. Wahyudiwan DDH, Sucahyo YG, Gandhi A. Information security awareness level measurement for employee: case study at Ministry of Research, Technology, and Higher Education. In: *Proceedings of the Third International Conference on Science in Information Technology*, Bandung, p. 25–6. 2017.
49. Shamsi AAA. Effectiveness of cyber security awareness program for young children: a case study in UAE. *Int J Inf Technol Lang Stud* 2019;**3**:8–29.
50. Gundu T, Flowerday S, Renaud K. Deliver security awareness training, then repeat: {Deliver, Measure Efficacy}. In: *Prroceedings of the Conference on Information Communications Technology and Society (ICTAS)*, Durban, March 6-8, 2019.
51. Ikhalia E, Serrano A, Bell D. *et al.* Online social network security awareness: mass interpersonal persuasion using a Facebook app. *Inf Technol People* 2019;**32**:1276–300.
52. Tschakert KF, Ngamsuriyaroj S. Effectiveness of and user preferences for security awareness training methodologies. *Heliyon* 2019;**5**:e02010.
53. Haney J, Lutters W. Security awareness training for the workforce: moving beyond "check-the-box" compliance. *Computer* 2020;**53**:91–5.
54. Parsons K, McCormac A, Pattinson M. *et al.* A study of information security awareness in Australian government organisations. *Inf Manag Comput Secur* 2014;**22**:334–45.
55. Kaur J, Mustafa N. Examining the effects of knowledge, attitude and behaviour on information security awareness: A case on SME. In: *Proceedings of the International Conference on Research and Innovation in Information Systems (ICRIIS)*, Kuala Lumpur, 2013.
56. Labuschagne W, Eloff M. The effectiveness of online gaming as part of a security awareness program. In: *Proceedings of the Thirteenth European Conference on Cyber Warfare and Security*, Piraeus, July 3–4, 2014.
57. Koyuncu M, Pusatli T. Security awareness level of smartphone users: an exploratory case study. *Mob Inf Syst* 2019;**2019**:1–11.
58. Ahlan AR, Lubis M, Lubis AR. Information security awareness at the knowledge-based institution: its antecedents and measures. *Proc Comput Sci* 2015;**72**:361–73.
59. Chen CC, Medlin BD, Shaw R. A cross-cultural investigation of situational information security awareness programs. *Inf Manag Comput Secur* 2008;**16**:360–76.
60. Eminağaoğlu M, Uçar E, Eren Ş. The positive outcomes of information security awareness training in companies: a case study. *Inf Secur Tech Rep* 2009;**14**:223–9.
61. Rantos K, Fysarakis K, Manifavas C. How effective is your security awareness program? An evaluation methodology. *Inf Secur J Glob Perspect* 2012;**21**:328–45.
62. Talib S, Clarke NL, Furnell SM. An analysis of information security awareness within home and work environments. In: *Proceedings of the International Conference on Availability, Reliability and Security*, Krakow, 2010.
63. Bulgurcu B, Cavusoglu H, Benbasat I. Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. *MIS Quart* 2010;**34**:523–48.
64. Kruse S, Pankey B. Assessing the effectiveness of security awareness training. http://www.securitymetrics.org/attachments/Metricon-6.5-Kruse.pdf. (20 September 2020, date last accessed).
65. Beyer M, Ahmed S, Doerlemann K. *et al.* Awareness is only the first step: a framework for progressive engagement of staff in cyber security. https://www.riscs.org.uk/wp-content/uploads/2015/12/Awareness-is-Only-the-First-Step.pdf. (20 September 2020, date last accessed).
66. Parsons K, McCormac A, Butavicius M. *et al.* Determining employee awareness using the Human Aspects of Information Security Questionnaire (HAIS-Q). *Comput Secur* 2014;**42**:165–76.
67. Egelman S, Peer E. Scaling the security wall: developing a security behaviour intention scale (SeBIS). In: *Proceedings of the Thirty-Third Annual ACM Conference on Human Factors in Computing Systems*, Seoul, April 18-23. 2015.
68. Faklaris C, Dabbish L, Hong JI. A self-report measure of end-user security attitudes (SA-6). In: *Proceedings of the USENIX Symposium on Usable Privacy and Security (SOUPS)*, Santa Clara, CA, August 11–13. 2019.
69. Rajivan P, Moriano P, Kelley T. *et al.* Factors in an end user security expertise instrument. *Inf Comput Secur* 2017;**25**:190–205.
70. Hadlington L. Human factors in cybersecurity, examining the link between Internet addiction, impulsivity, attitudes towards cybersecurity, and risky cybersecurity behaviours. *Heliyon* 2017;**3**:e00346.
71. Coutlee CG, Politzer CS, Hoyle RH. *et al.* An abbreviated impulsiveness scale (ABIS) constructed through confirmatory factor analysis of the BIS-11. *Arch Sci Psychol* 2014;**2**:1–12.
72. Davis RA, Flett GL, Besser A. Validation of a new scale for measuring problematic internet use: implications for pre-employment screening. *Cyberpsychol Behav* 2002;**5**:331–45.
73. Ögăütçü G, Testik ÖM, Chouseinoglou O. Analysis of personal information security behavior and awareness. *Comput Secur* 2016;**56**:83–93.
74. Huang HY, Demetriou S, Banerjee R. *et al.* Smartphone security behavioral scale: a new psychometric measurement for smartphone security. https://arxiv.org/abs/2007.01721. (26 August 2020, date last accessed).
75. Velki T, Solic K, Ocevcic H. Development of Users' Information Security Awareness Questionnaire (UISAQ)—ongoing work. In: *Proceedings of the Thirty-Seventh International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, May 26–30. 2014.
76. Caballero A. Security education, training, and awareness. In: *Computer and Information Security Handbook*, Burlington: Morgan Kaufmann Publishers, 2017, 497–505.
77. Spitzner L. Security awareness for senior management. (29 July 2021, date last accessed).
78. Peltier TR. Implementing an information security awareness program. *Inf Syst Secur* 2005;**14**:37–49.
79. Abawajy J. User preference of cyber security awareness delivery methods. *Behav Inf Technol* 2014;**33**:237–48.
80. Nachin N, Tangmanee C, Piromsopa K. How to increase cybersecurity awareness. *ISACA J* 2019;**2**:45–50.
81. González CSG, Toledo P, Izquierdo FB. Integrating the principles of DGBL, CSCL and playability in the design of social videogames: a case of study. In: *Student Usability in Educational Software and Games: Improving Experiences*, Hershey, IGI Global, 2012, 293–304.
82. Mabitle K, Kritzinger E. Schoolteacher preference of cyber-safety awareness delivery methods: a South African study. In: Silhavy R. (eds), *Artificial Intelligence and Bioinspired Computational Methods. CSOC 2020. Advances in Intelligent Systems and Computing*, Cham: Springer, 2020.
83. Ainley M, Hidi S, Berndroff D. Interest, learning, and the psychological processes that mediate their relationship. *J Educ Psychol* 2002;**94**:545–61.
84. Rounds J, Su R. The nature and power of interests. *Curr Dir Psychol Sci* 2014;**23**:98–103.
85. Shen X, Eades P, Hong S. *et al.* Intrusive and non-intrusive evaluation of ambient displays. In: *Proceedings of the First International Workshop on Ambient Information Systems, Collocated at Pervasive*, Toronto, 2007.
86. Volkamer M, Sasse MA, Boehm F. Analysing simulated phishing campaigns for staff. In: *Proceedings of the ESORICS Second Workshop on Security, Privacy, Organizations, and Systems*, Guildford, 2020.
87. Dupuis MJ, Smith S. Clickthrough testing for real-world phishing simulations. In: *Proceedings of the Twenty-First Annual Conference on Information Technology Education*, Online event, 2020.
88. Mustaca S Define S.M.A.R.T IT Security Goals. (ISC)2. https://blog.isc2.org/isc2_blog/2013/02/define-smart-it-security-goals.html. (23 March 2021, date last accessed), 2013.
89. Ceneric I, Looney J, Greef Md. *Indicators for Evaluation of Awareness and Fundraising for Low Literacy in Europe*. Brussels: ELINET- European Literacy Policy Network, 2014.
90. BetterEvaluation. Define ethical and quality evaluation standards. https://www.betterevaluation.org/en/rainbow_framework/manage/define_ethical_and_quality_evaluation_standards. (23 July 2012, date last accessed).