Train as you Fight: Evaluating Authentic Cybersecurity Training in Cyber Ranges

Magdalena Glas
University of Regensburg
Regensburg, Germany
magdalena.glas@ur.de

Manfred Vielberth
University of Regensburg
Regensburg, Germany
manfred.vielberth@ur.de

Günther Pernul
University of Regensburg
Regensburg, Germany
guenther.pernul@ur.de

ABSTRACT

Humans can play a decisive role in detecting and mitigating cyber attacks if they possess sufficient cybersecurity skills and knowledge. Realizing this potential requires effective cybersecurity training. Cyber range exercises (CRXs) represent a novel form of cybersecurity training in which trainees can experience realistic cyber attacks in authentic environments. Although evaluation is undeniably essential for any learning environment, it has been widely neglected in CRX research. Addressing this issue, we propose a taxonomy-based framework to facilitate a comprehensive and structured evaluation of CRXs. To demonstrate the applicability and potential of the framework, we instantiate it to evaluate Iceberg CRX, a training we recently developed to improve cybersecurity education at our university. For this matter, we conducted a user study with 50 students to identify both strengths and weaknesses of the CRX.

CCS CONCEPTS

• **Applied computing → Interactive learning environments**; • **Security and privacy → Social aspects of security and privacy**.

KEYWORDS

cyber range, evaluation method, cybersecurity exercise, cyber defense exercise

ACM Reference Format:
Magdalena Glas, Manfred Vielberth, and Günther Pernul. 2023. Train as you Fight: Evaluating Authentic Cybersecurity Training in Cyber Ranges. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3544548.3581046

1 INTRODUCTION

Humans have the ability to observe and interpret potentially malicious activities that indicate cyberattacks beyond what purely technical security measures can detect. This way, people can make a valuable contribution to protecting themselves and the environment in which they work or live from the risks of cyberattacks [62]. This realization can be seen as a paradigm shift [88], as humans are no longer considered the "weakest link" [27, 43, 65] in cybersecurity but a "source of strength" [18] for detecting sophisticated attacks. To empower humans to realize this potential, it is imperative to equip them with sufficient skills and knowledge in the field of cybersecurity. In recent years, there has been a noticeable trend that cybersecurity training is more effective if it is not based on the conventional transfer of theoretical knowledge in the form lectures or static e-learning but rather on the interactive transfer of hands-on skills in authentic environments [14, 83]. Cybersecurity exercises in cyber ranges are a common concept to realize this approach. Cyber ranges are realistic simulations or emulations of real networks or applications that provide an environment for cybersecurity training [48]. Taking part in a cyber range exercise (CRX), trainees can expose themselves to realistic security-relevant situations to train as though they were fighting cyber attacks in the real world. As with any training or learning intervention, evaluating the effectiveness and efficiency of a CRX is clearly desirable. Evaluation results provide a source for understanding what aspects of a CRX need to be improved to achieve the desired training outcome with assigned resources. Therefore, evaluation is a fundamental requirement to ensure the quality of a CRX and allow CRX developers to continuously improve their product. It is unlikely that a CRX whose effectiveness has not been validated will be implemented in an academic or organizational cybersecurity training program. The critical role of evaluation in CRX design is also recognized in the academic literature where evaluation is identified as an essential step of a CRX's life-cycle [38, 79, 83]. In recent years, a plethora of CRXs has been introduced in the literature [72, 83]. While these concepts are often described in great technical detail, an evaluation of the CRX is rarely provided. With our research we seek to facilitate CRX evaluation to enable CRXs to reach their full potential. This leads us to the following simple but encompassing research question:

RQ. How can CRXs be evaluated?

Contribution. We address this research question by proposing the *TARGET* framework, which aims to support CRX designers in research and practice in evaluating CRXs in a structured manner. We argue it is desirable to design a framework that is universally applicable to a wide range of CRX concepts. This way, the framework not only enables CRX designers to continuously improve their products over time but also provides comparability between different CRXs. Comparable evaluation designs enable organizations to make better-informed decisions about which existing CRX fits their learners' needs best. This way, not only CRX designers but also learners, e.g., students at a university or employees at a company, benefit from a universal evaluation approach. For an evaluation framework to provide this universality, we examine the existing literature to develop a taxonomy for CRX evaluation criteria and

CHI '23, April 23–28, 2023, Hamburg, Germany
© 2023 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3581046

embed the taxonomy in a five-step evaluation process. Together, the *TARGET* taxonomy and the *TARGET* process build the *TARGET* framework. To demonstrate the framework's applicability, we use it for evaluating Iceberg CRX, a CRX we recently developed to improve cybersecurity education at our university. In Iceberg CRX, trainees take on the role of incident responders in a Security Operations Center (SOC) and learn to detect and respond to cybersecurity attacks against a simulated industrial system. Using the *TARGET* framework, we conducted a full evaluation of Iceberg CRX, the central element of which was a user study with 50 participants. The evaluation results highlight the potential Iceberg CRX provides to improve cybersecurity education but also help to identify which aspects of the CRX do not yet meet the set evaluation criteria and thus point the way to future improvements.

2 CYBER RANGE EXERCISES

The concept of cyber ranges emerged from the military sector, where the first cyber range was developed in 2008 by the US department of defense, conceptualized as a form of shooting range for cybersecurity. Today, cyber ranges are not only used in the military field but have a variety of application purposes. The National Institute of Standardization (NIST) defines cyber ranges as "interactive, simulated representations of an organization's local network, system, tools, and applications that are connected to a simulated Internet level environment" [48] and provide a safe, legal and secure environment for security education, training, and testing. This environment can be entirely virtual but might also include actual physical hardware [35].

In this work, the concept of cyber ranges is examined in terms of their purpose for cybersecurity training through CRXs. In a CRX, a trainee can acquire cybersecurity skills and knowledge through a hands-on experience in an environment that closely resembles reality. Here, skills and knowledge can be imparted to the trainee via two perspectives common to cybersecurity exercises. In a defense-oriented approach (*Blue Teaming*), trainees learn how attacks can be detected and mitigated. In an attack-oriented approach (*Red Teaming*), trainees take an attacker's perspective and learn how a system's vulnerabilities can be discovered and exploited. Through the latter, the trainee learns to understand which vulnerabilities in a system offer potential entry points for an attacker and, conversely, how a system should be designed to be more resilient against cyber attacks.

To give an overview of the capabilities of cyber ranges, we refer to the work of Yamin et al. [83]. The authors propose a taxonomy for classifying cyber range concepts composed of six capabilities, which we describe below. In the following, the term *capability* is used in the sense of their classification. **Capabilities of cyber ranges** (based on Yamin et al. [83]):

- **Scenario:** A scenario encompasses the context and storyline of a CRX, including its target group and the domain in which the cyber range resides, e.g., critical systems or industrial IoT.
- **Environment:** The environment is the technical implementation of a CRX's scenario. This includes the software and hardware used to simulate or emulate a system and the attacks against it.
- **Teaming:** The teaming aspect of a cyber range describes which roles exist in a CRX, who takes on these roles, or to what extent they are automated. Besides the Blue and Red Team, a White Team, responsible for managing the infrastructure and exercise, and a Green Team, responsible for maintaining the infrastructure, is often specified.
- **Learning:** Learning encompasses all aspects of the CRX related to the transfer of knowledge and skills. Learning is commonly realized through a Learning Management System (LMS), which typically entails tasks for the trainees to solve throughout the CRX combined with a gamified scoring system as well as tutoring elements, e.g., in the form of texts and visual elements, that support the trainee in understanding the scenario.
- **Monitoring:** This capability describes which processes and mechanisms are used to monitor both the technical infrastructure and the trainees' behavior during the exercise.
- **Management:** Management comprises all aspects of organizing a CRX and managing the cyber range's infrastructure.

3 RELATED WORK

To frame our research, we examine prior works from two angles. In Section 5, we analyze how CRX concepts have been evaluated in the literature (without a comprehensive CRX evaluation framework available). In this section, we contextualize our work within related evaluation frameworks. These frameworks originate from the research area of cybersecurity exercises, of which CRXs represent a subset. We discuss the suitability of these frameworks for CRX evaluation and outline how the proposed *TARGET* framework goes beyond what has been achieved so far.

First, we identified frameworks for evaluating trainees' learning in a cybersecurity exercise. Patriciu et al. [52] present a sequential process for evaluating cybersecurity exercises based on the definition of learning objectives. The design and evaluation of the exercise should be aligned with these learning objectives. Dark et al. [20] apply evaluation practices for educational exercises to the cybersecurity domain and propose a five-step process for evaluating to which extent a cybersecurity exercise succeeds in fulfilling its purpose. Similar to the approach by Patriciu et al. [52], the framework proposes to base the evaluation process upon predefined goals. The authors state that these goals might be both short-term learning objectives and long-term goals, e.g., improving trainees' security mindsets in their daily lives. To assess the suitability of these frameworks for evaluating CRX, we set them in the context of Yamin et al.'s [83] cyber range capabilities described above. While the two frameworks provide comprehensive insights into how *Learning* can be evaluated, other capabilities, such as the technical implementation (*Environment*) or the organization of a CRX (*Management*), are not considered. For this matter, we propose the *TARGET* framework, which adapts the goal-based evaluation approach for assessing aspects across all capabilities of a CRX. In contrast to the two frameworks above that only mention a few examples for evaluation goals, our framework provides a comprehensive knowledge base for evaluation goals in the form of a literature-based taxonomy. This taxonomy allows CRX designers to identify relevant goals for their CRX and conduct an evaluation accordingly.

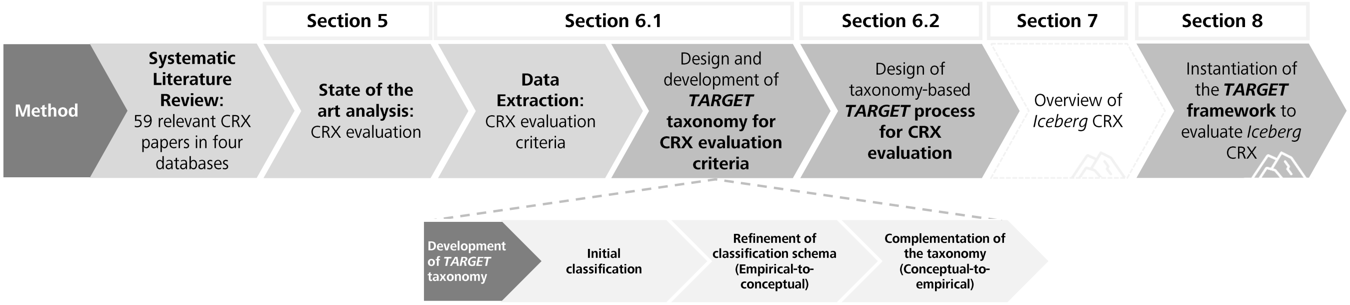

Section 5
Section 6.1
Section 6.2
Section 7
Section 8
Method
Systematic Literature Review: 59 relevant CRX papers in four databases
State of the art analysis: CRX evaluation
Data Extraction: CRX evaluation criteria
Design and development of TARGET taxonomy for CRX evaluation criteria
Design of taxonomy-based TARGET process for CRX evaluation
Overview of Iceberg CRX
Instantiation of the TARGET framework to evaluate Iceberg CRX
Development of TARGET taxonomy
Initial classification
Refinement of classification schema (Empirical-to-conceptual)
Complementation of the taxonomy (Conceptual-to-empirical)

Figure 1: Methodical approach for designing *TARGET* framework: The method we follow in this paper follows seven steps, each corresponding to one section or subsection of the paper. Step 4, the taxonomy again consists of three sub-steps that follow the taxonomy design method of Kundisch et al. [39].

Furthermore, we identified a framework by Chowdhury et al. [16] for the overall design, development, and execution of cybersecurity exercises that translates well to CRXs. The framework is based on the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) model and acknowledges evaluation as an integral part of a cybersecurity exercise's continuous design and development process. The authors propose to evaluate not only *Learning* but also other factors, such as the quality of the implementation of an exercise. However, they provide little guidance on selecting these metrics to design to conduct an evaluation. In contrast, the *TARGET* framework provides CRX designers with a fine-grained process for selecting and assessing evaluation metrics in a structured manner.

Finally, several works present frameworks for evaluating specific aspects of a cybersecurity exercise. These frameworks, however, are no generic methods like the ones described above but practical evaluation approaches in the form of tools facilitating evaluation [2, 10, 44, 78] or specific evaluation instruments (e.g., scales) [11, 30, 40]. Andreolini et al. [2], Braghin et al. [10] and Švábenský et al. [78] propose frameworks for monitoring and evaluating trainees' performance by creating directed graphs that visualize trainees' activities during an exercise. Maennel et al. [44] develop a timestamp-based framework to measure the trainees' learning process during an exercise based on log data and interview-based surveys. Granåsen and Andersson [30], Buchler et al. [11] and Fleur et al. [40] present frameworks for assessing team effectiveness in a cybersecurity exercise. Thus, not only the extent to which a team of trainees achieves a given task but also how team members collaborate in the process and which impact certain teaming qualities have on the learning outcome of a team. While these frameworks aim to improve the evaluation of particular aspects of a CRX, the *TARGET* goes one step further and intends to assess whether a CRX fulfills its purpose from a global perspective. To this end, it enables the flexible orchestrating of different evaluation tools and instruments based on predefined goals.

In essence, due to the socio-technical complexity of a CRX, no existing evaluation framework is adequate to evaluate a CRX in its entirety. This is evidenced by the fact that, to date, CRX evaluation is mostly conducted in an unstructured manner without referencing an evaluation framework – a matter we will discuss in detail in Section 5. With *TARGET*, we pursue to overcome this issue by providing a structured evaluation process that allows evaluating aspects across all capabilities of a CRX – an endeavor that, to the best of our knowledge, has not been attempted yet.

4 METHOD

The research methodology followed in this paper consists of seven steps, which are illustrated in Figure 1 and outlined below. For our framework to be universally applicable to a variety of CRXs, it is necessary to build it upon contemporary CRX research. For this reason, we performed a structured literature review (**Step 1**) on CRXs. To collect a representative corpus of related papers, we performed a systematic search in the four online libraries IEEE Xplore, ACM Digital Library, SpringerLink, and ScienceDirect, using the search terms "cyber range* AND training" and "cyber range* AND exercise*" in title, keywords and abstract of the papers. The search was conducted in two iterations (April 2022 and November 2022) and resulted in a total of 113 papers. Each paper was individually screened regarding rigor and its relevance for our research to remove papers that did not present a CRX concept but, e.g., CRX-related literature reviews or CRX design frameworks. After this step, we obtained a total of 59 papers about CRX concepts from the years of 2016 to 2022[1]. To investigate the issue of missing evaluation in CRX research more precisely, we examined the current state of the art of CRX evaluation (**Step 2**). We analyzed which of the 59 includes an evaluation and, if so, which CRX capabilities are covered by the evaluation. This analysis, whose results are outlined in detail in Section 5, showed that CRX evaluation is either neglected entirely or carried out in an unstructured manner.

We seek to overcome this issue with a framework that allows evaluating a CRX in a comprehensive and structured way. Because of the diversity of CRXs, the objectives and requirements of a CRX, and thus which aspects are subject to an evaluation, can vary widely [16]. For a large-scale CRX involving hundreds of participants at a time, different aspects will be of interest for evaluation than for a CRX tailored to an organization's needs and a small target group of highly-skilled experts. For our framework to provide this flexibility, we propose to use a goal-based evaluation approach. Identifying goals as a basis for evaluation is a common approach to instructional design [59], but also software development [58], both of which CRX

[1] For traceability, a complete list of all papers and our filtering steps is published on GitHub: https://github.com/TARGETframework/LiteratureReview

Table 1: Capabilities covered by the 31 CRX concepts that included an evaluation: the capabilities mostly covered by CRX evaluation in the analyzed corpus of literature are *Learning* (17 papers) and *Environment* (15 papers). *Teaming, Scenario, Management*, and *Monitoring* were fewer subject to CRX evaluation.

Author	Year	Scenario	Environment	Teaming	Learning	Monitoring	Management
Arshad et al. [4]	2021				●		
Bernardinetti et al. [6]	2021						●
Beuran et al. [8]	2018	●	●				●
Beuran et al. [7]	2019	●	●				●
Beuran et al. [9]	2022		●				
Caturano et al. [13]	2020	●					
Caturano et al. [12]	2022		●				
Fenton et al. [25]	2019				●		
Glas et al. [29]	2022				●		
Hatzivasilis et al. [32]	2021	●	●			●	
Jacq et al. [33]	2021				●		
Nakata and Otsuka [47]	2021		●				
Ošlejšek et al. [51]	2018		●		●		
Peratikou et al. [53]	2021		●				
Pham et al. [54]	2016	●	●				
Puys and Mocanu [57]	2021				●		
Raybourn et al. [60]	2018				●		
Roberts et al. [63]	2021	●					
Russo et al. [64]	2020		●				
Shrivastava et al. [67]	2022		●				
Švábenský et al. [77]	2018			●	●		
Tang et al. [70]	2017		●		●		
Tian et al. [71]	2018					●	
Vekaria et al. [74]	2021		●		●		
Vielberth et al. [76]	2021				●		
Vykopal et al. [79]	2017				●		
Yamin et al. [84]	2021				●		
Yamin et al. [82]	2022		●		●		●
Yamin et al. [81]	2022		●		●		●
Yonemura et al. [85]	2021				●		
Yonemura et al. [86]	2022				●		
	Σ	6	15	1	17	2	5

design intersects with. To this end, we recommend defining evaluation criteria, each of which relates to a goal pursued by the design of the CRX. For each criterion, one or more metrics are then defined to measure its fulfillment. This approach allows for the evaluation of a CRX in a targeted manner while following a structured and transparent evaluation process. To enable CRX designers to make an informed decision about which evaluation criteria are relevant to their CRX, we developed a literature-based taxonomy of evaluation criteria for CRXs (**Step 3**). To extract possible evaluation criteria from the literature, we examined what requirements and objectives authors define for their CRX. To classify these criteria, we followed the taxonomy design method of Kundisch et al. [39] and performed three steps (1) Initial classification, (2) Refinement of the classification scheme, and (3) Complementation of the taxonomy (**Step 4**). These steps are described in detail in Section 6.1. Subsequently, we present the *TARGET* process in Section 6.2 (**Step 5**), in which we propose how to utilize the taxonomy to design and conduct a CRX evaluation. After presenting the concept and implementation of Iceberg CRX in Section 7 (**Step 6**), we demonstrate the application of the framework for its evaluation in Section 8 (**Step 7**).

5 EXPLORING THE NEED FOR STRUCTURED CRX EVALUATION

Evaluation can provide CRX designers with information on the extent to which the CRX fulfills its purpose and which aspects of the CRX need to be improved. In addition, the results of an evaluation give cybersecurity researchers, educators, and learners important information about a CRX's quality. Out of the 59 identified CRX papers, only 31 included an evaluation. This highlights that evaluation, while undoubtedly important, is still largely neglected in CRX development. This limits the practical use of a CRX and, thus, the organizational and societal value it could potentially provide.

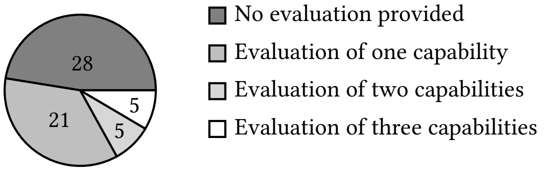


Figure 2: Results of the analysis: 31 of 59 CRXs include an evaluation out of which 21 papers cover one capability, 5 papers cover two capabilities, and 5 papers cover three capabilities in the respective evaluation.

The 31 papers that covered evaluation were classified for the capabilities subject to the evaluation. The result of this analysis is presented in Table 1 and Figure 2. While some authors evaluate up to three capabilities of the CRX, most works merely consider one capability. The latter is often the case when a particular aspect of one of the capabilities is being investigated or improved. For

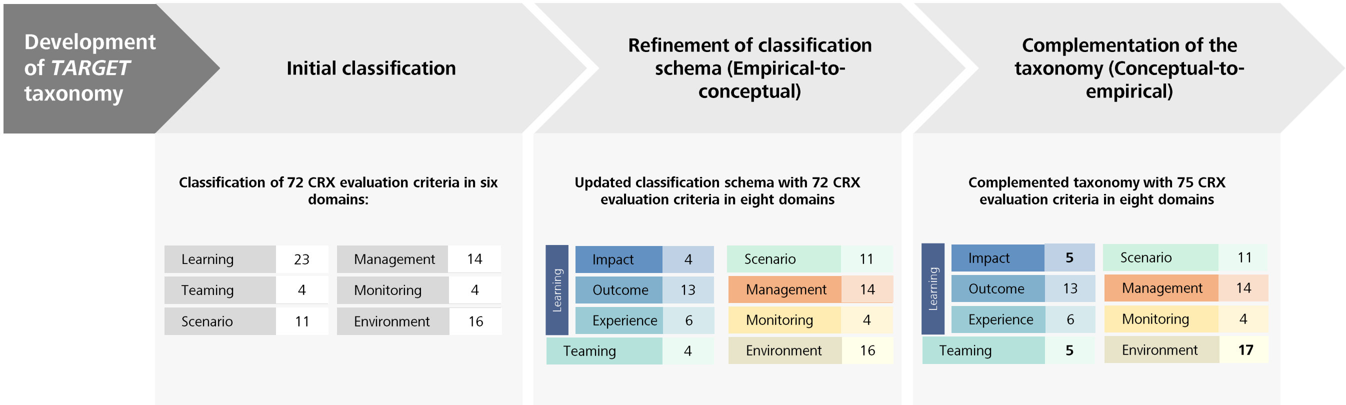

Development of *TARGET* taxonomy
Initial classification
Refinement of classification schema (Empirical-to-conceptual)
Complementation of the taxonomy (Conceptual-to-empirical)
Classification of 72 CRX evaluation criteria in six domains:
Learning 23
Teaming 4
Scenario 11
Management 14
Monitoring 4
Environment 16
Updated classification schema with 72 CRX evaluation criteria in eight domains
Learning
Impact 4
Outcome 13
Experience 6
Teaming 4
Scenario 11
Management 14
Monitoring 4
Environment 16
Complemented taxonomy with 75 CRX evaluation criteria in eight domains
Learning
Impact 5
Outcome 13
Experience 6
Teaming 5
Scenario 11
Management 14
Monitoring 4
Environment 17

Figure 3: Process followed for developing the taxonomy: the 72 criteria extracted from the literature review were initially classified using the capabilities by Yamin et al. [83] before refining this classification and complementing the taxonomy with three further criteria in the dimensions *Learning Impact* and *Environment*.

example, some CRXs are primarily concerned with improving the learning process and content [4, 60, 76, 79, 85], while others focus on specific aspects of the CRX's environment and thus primarily evaluate technical aspects such as the performance of the proposed system [47, 53, 64]. Concerning the evaluation design of the CRXs under review, it can be stated that there is no standard approach the evaluation follows. Regarding the design of the evaluation, we found that in none of the cases did the authors follow an established methodology or framework like the ones discussed in Section 3.

6 THE *TARGET* FRAMEWORK

We address the identified need for more comprehensive and better structured CRX evaluation by proposing *TARGET*, a framework that aims to both standardize and facilitate CRX evaluation. *TARGET* intends support CRX designers in research and practice in conducting a comprehensive summative evaluation of their product. Thereby, it acknowledges the diversity of different CRXs and provides an evaluation method that is adaptable to the specific needs of different CRXs.

6.1 The *TARGET* Taxonomy: Classifying CRX evaluation criteria

The taxonomy was developed in three steps following the method by Kundisch et al. [39], which we introduced in Section 4. These steps which represent a refinement of Step 4 of this work's method (rf. Section 4) are illustrated in Figure 3 and outlined in the following.

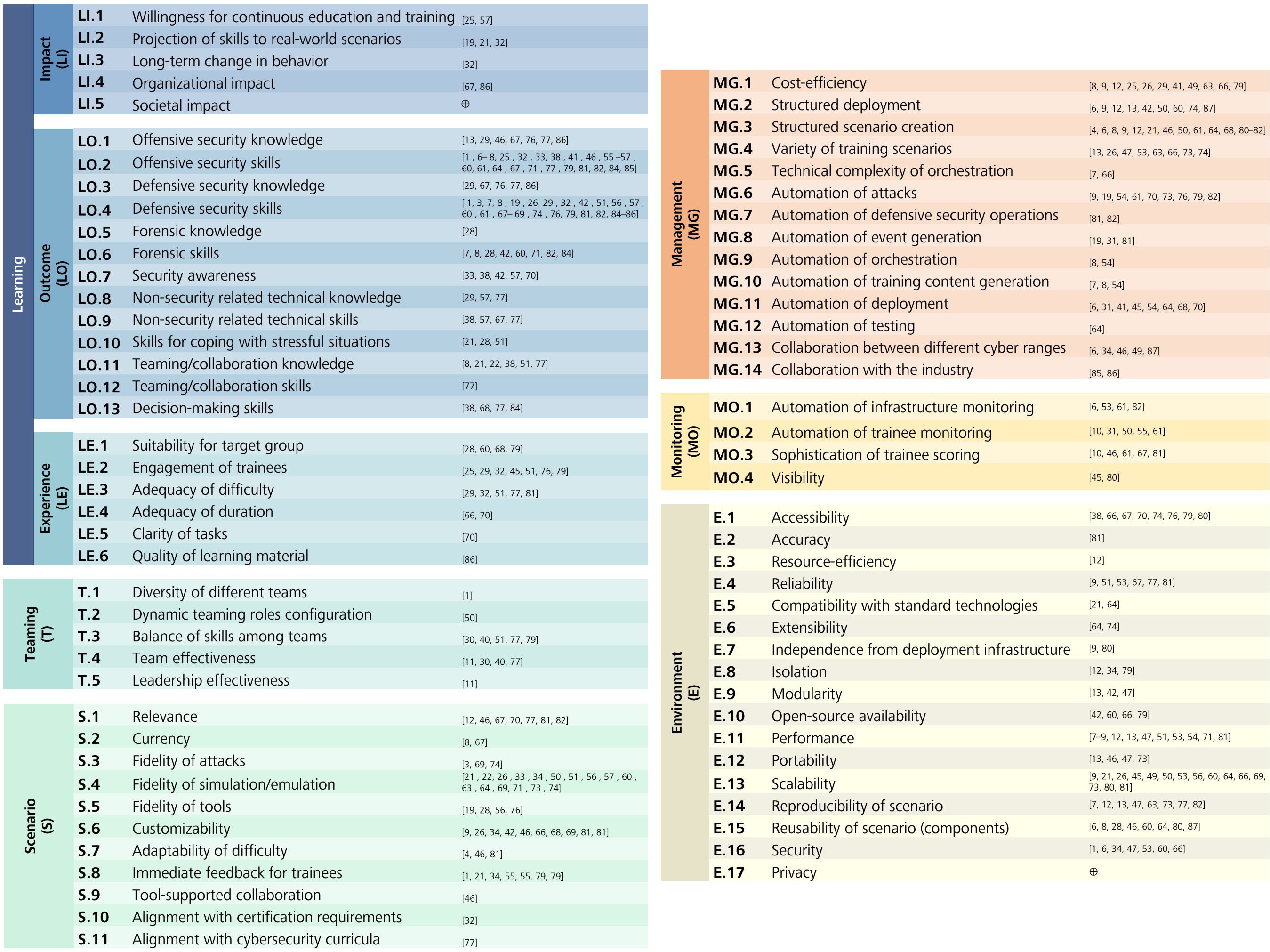
(1) **Initial Classification (empirical-to-conceptual):** We could extract 72 different evaluation criteria from the 59 papers included in our literature review. In the course of an initial conceptualization, we utilized the capabilities by Yamin et al. [83] (*Scenario, Environment, Teaming, Learning, Monitoring* and *Management*) for classifying the criteria.

(2) **Refinement of Classification:** Subsequently, the initial classification was refined. Since learning should be the ultimate goal for any CRX, we decided to address this aspect of evaluation more precisely. To do so, we adopted Kirkpatrick's [37] model for training evaluation. The model perceives training evaluation to happen on four levels: the learning process itself (Level 1), the extent to which trainees improve their skills and knowledge (Level 2), how they adapt that knowledge to their everyday work (Level 3), and the organizational impact of that behavior change (Level 4). To simplify the model's application for *TARGET*, we combine the last two levels, which results in the three dimensions *Learning Experience* (Level 1), *Learning Outcome* (Level 2), and *Learning Impact* (Level 3 and Level 4), leading to overall eight dimensions for classification:

- **Learning Impact (LI):** Evaluation criteria related to the long-term effects the CRX pursues to achieve.
- **Learning Outcome (LO):** Evaluation criteria related to the skills and knowledge that are conveyed through the CRX.
- **Learning Experience (LE):** Evaluation criteria related to the the trainees' response to the learning process.
- **Teaming (T):** Evaluation criteria related to the organization and trainee staffing of the teams within the CRX.
- **Scenario (S):** Evaluation criteria related to the storyline, context, and domain of the CRX.
- **Management (MG):** Evaluation criteria related to the preparation and execution of the CRX.
- **Monitoring (MO):** Evaluation criteria related to the modalities of overseeing the operation of the CRX and collecting and analyzing data to assess the fulfillment of goals for evaluation.
- **Environment (E):** Evaluation criteria related to the technical implementation of the CRX.

We then specified each criterion by its name and an ID containing the abbreviation of the dimension to which the criterion was classified, e.g., **Suitability for target group (LE.1)**.

(3) **Complementation of the Taxonomy (conceptual-to-empirical):** In a final step, we complemented the taxonomy



Learning	Impact (LI)	LI.1	Willingness for continuous education and training	[25, 57]
		LI.2	Projection of skills to real-world scenarios	[19, 21, 32]
		LI.3	Long-term change in behavior	[32]
		LI.4	Organizational impact	[67, 86]
		LI.5	Societal impact	⊕
	Outcome (LO)	LO.1	Offensive security knowledge	[13, 29, 46, 67, 76, 77, 86]
		LO.2	Offensive security skills	[1 , 6– 8, 25 , 32 , 33, 38 , 41 , 46 , 55 –57 , 60, 61, 64 , 67 , 71 , 77 , 79, 81, 82, 84, 85]
		LO.3	Defensive security knowledge	[29, 67, 76, 77, 86]
		LO.4	Defensive security skills	[1, 3, 7, 8 , 19 , 26, 29 , 32 , 42 , 51, 56 , 57 , 60 , 61 , 67– 69 , 74 , 76, 79, 81, 82, 84–86]
		LO.5	Forensic knowledge	[28]
		LO.6	Forensic skills	[7, 8, 28, 42, 60, 71, 82, 84]
		LO.7	Security awareness	[33, 38, 42, 57, 70]
		LO.8	Non-security related technical knowledge	[29, 57, 77]
		LO.9	Non-security related technical skills	[38, 57, 67, 77]
		LO.10	Skills for coping with stressful situations	[21, 28, 51]
		LO.11	Teaming/collaboration knowledge	[8, 21, 22, 38, 51, 77]
		LO.12	Teaming/collaboration skills	[77]
		LO.13	Decision-making skills	[38, 68, 77, 84]
	Experience (LE)	LE.1	Suitability for target group	[28, 60, 68, 79]
		LE.2	Engagement of trainees	[25, 29, 32, 45, 51, 76, 79]
		LE.3	Adequacy of difficulty	[29, 32, 51, 77, 81]
		LE.4	Adequacy of duration	[66, 70]
		LE.5	Clarity of tasks	[70]
		LE.6	Quality of learning material	[86]
Teaming (T)		T.1	Diversity of different teams	[1]
		T.2	Dynamic teaming roles configuration	[50]
		T.3	Balance of skills among teams	[30, 40, 51, 77, 79]
		T.4	Team effectiveness	[11, 30, 40, 77]
		T.5	Leadership effectiveness	[11]
Scenario (S)		S.1	Relevance	[12, 46, 67, 70, 77, 81, 82]
		S.2	Currency	[8, 67]
		S.3	Fidelity of attacks	[3, 69, 74]
		S.4	Fidelity of simulation/emulation	[21 , 22, 26 , 33 , 34 , 50 , 51 , 56 , 57 , 60 , 63 , 64 , 69, 71 , 73 , 74]
		S.5	Fidelity of tools	[19, 28, 56, 76]
		S.6	Customizability	[9, 26, 34, 42, 46, 66, 68, 69, 81, 81]
		S.7	Adaptability of difficulty	[4, 46, 81]
		S.8	Immediate feedback for trainees	[1, 21, 34, 55, 55, 79, 79]
		S.9	Tool-supported collaboration	[46]
		S.10	Alignment with certification requirements	[32]
		S.11	Alignment with cybersecurity curricula	[77]

Management (MG)	MG.1	Cost-efficiency	[8, 9, 12, 25, 26, 29, 41, 49, 63, 66, 79]
	MG.2	Structured deployment	[6, 9, 12, 13, 42, 50, 60, 74, 87]
	MG.3	Structured scenario creation	[4, 6, 8, 9, 12, 21, 46, 50, 61, 64, 68, 80–82]
	MG.4	Variety of training scenarios	[13, 26, 47, 53, 63, 66, 73, 74]
	MG.5	Technical complexity of orchestration	[7, 66]
	MG.6	Automation of attacks	[9, 19, 54, 61, 70, 73, 76, 79, 82]
	MG.7	Automation of defensive security operations	[81, 82]
	MG.8	Automation of event generation	[19, 31, 81]
	MG.9	Automation of orchestration	[8, 54]
	MG.10	Automation of training content generation	[7, 8, 54]
	MG.11	Automation of deployment	[6, 31, 41, 45, 54, 64, 68, 70]
	MG.12	Automation of testing	[64]
	MG.13	Collaboration between different cyber ranges	[6, 34, 46, 49, 87]
	MG.14	Collaboration with the industry	[85, 86]
Monitoring (MO)	MO.1	Automation of infrastructure monitoring	[6, 53, 61, 82]
	MO.2	Automation of trainee monitoring	[10, 31, 50, 55, 61]
	MO.3	Sophistication of trainee scoring	[10, 46, 61, 67, 81]
	MO.4	Visibility	[45, 80]
Environment (E)	E.1	Accessibility	[38, 66, 67, 70, 74, 76, 79, 80]
	E.2	Accuracy	[81]
	E.3	Resource-efficiency	[12]
	E.4	Reliability	[9, 51, 53, 67, 77, 81]
	E.5	Compatibility with standard technologies	[21, 64]
	E.6	Extensibility	[64, 74]
	E.7	Independence from deployment infrastructure	[9, 80]
	E.8	Isolation	[12, 34, 79]
	E.9	Modularity	[13, 42, 47]
	E.10	Open-source availability	[42, 60, 66, 79]
	E.11	Performance	[7–9, 12, 13, 47, 51, 53, 54, 71, 81]
	E.12	Portability	[13, 46, 47, 73]
	E.13	Scalability	[9, 21, 26, 45, 49, 50, 53, 56, 60, 64, 66, 69, 73, 80, 81]
	E.14	Reproducibility of scenario	[7, 12, 13, 47, 63, 73, 77, 82]
	E.15	Reusability of scenario (components)	[6, 8, 28, 46, 60, 64, 80, 87]
	E.16	Security	[1, 6, 34, 47, 53, 60, 66]
	E.17	Privacy	⊕

Figure 4: The *TARGET* taxonomy of CRX evaluation criteria: the taxonomy we present consists of 75 evaluation criteria, classified into eight dimensions. For full accessibility, we would like to refer to the appendix of this paper, in which we present the taxonomy with one table for each dimension (rf. Tables 5-12).

with evaluation criteria that do not appear in the identified CRX literature yet contribute to the completeness of the taxonomy. This estimation is based on the insights we gained from developing the classification scheme and the expertise we gained from developing and conducting CRXs ourselves. Examining the literature, it was striking that very few papers mention the long-term effects the CRX seeks to achieve (*Learning Impact*). Although these are the evaluation criteria of CRX that are certainly the most difficult to assess due to their long-term focus, they represent overarching goals that the entire CRX should be aligned with. To represent this dimension comprehensively, we complemented with the evaluation criterion **Societal Impact (LI.5)**. Additionally, we complemented the dimension *Environment* with the goal **Privacy (E.17)** as in a CRX, personal data of the trainees may be obtained, and therefore, naturally, their privacy must be protected. In the taxonomy, these complemented criteria are labeled with the symbol ⊕ (rf. Figure 4).

Teaming has received little attention in the CRX literature, as evidenced by the low number of criteria in this dimension after the initial extraction. However, several researchers [11, 30, 40] have extensively studied team behavior in cybersecurity exercises (rf. Section 3). Since we believe that their findings are equally applicable to CRXs and represent a useful complement to the *Teaming* dimension, we included these works into the *Teaming* dimension of the taxonomy. This way, we were able to add the criterion **Leadership effectiveness (T.5)** and provide a more comprehensive set of references for the criteria **Balanced of skills among teams (T.3)** and **Team effectiveness (T.4)**.

In Figure 4, we present the complete *TARGET* taxonomy with overall 75 CRX evaluation criteria classified into eight dimensions.

As is the nature of a taxonomy [39], this reflects the state of CRX evaluation criteria at the present moment and builds the foundation for dynamic extensions and adjustments to the taxonomy in the future.

6.2 *TARGET* Process: Providing a structured CRX evaluation design

Building on the taxonomy described in the previous subsection, we now present a five-step process that describes how to perform a complete CRX evaluation (rf. Figure 5). While the *TARGET* taxonomy provides the knowledge base to build the evaluation upon contemporary CRX research, the associated *TARGET* process specifies the structure and sequence of the evaluation.

(1) **Selection of Evaluation Criteria:** The first step of the process is the selection of those evaluation criteria from the *TARGET* taxonomy that are relevant to the CRX. For each criterion, it needs to be precisely determined how it is to be addressed by the CRX. The sequence in which the evaluation criteria are selected follows a top-down approach (rf. Figure 5). As outlined before, learning represents the overarching goal of a CRX. Hence, one should first consider the long-term impact that should be achieved by the CRX (*Learning Impact*), what skills and knowledge are necessary for the trainees to learn as a basis for this impact (*Learning Outcome*) and how the learning process should be designed to achieve the desired outcome (*Learning Experience* and *Teaming*). The basis for this is laid by the design of the scenario of the CRX, including the storyline of the exercise (*Scenario*). This translates into criteria for the CRX's organization and governance (*Management* and *Monitoring*) and, finally, the criteria of the technical implementation of the CRX (*Environment*).

(2) **Selection of Metrics:** Subsequently, one or more metrics are assigned to each criterion. For this step, we would like to refer to the work of Basili [5], which proposes a paradigm for the structured selection of evaluation metrics. Along with the literature review we conducted on CRX evaluation (rf. Section 5), we extracted metrics used for CRX evaluation. This set of metrics, however, does not cover all evaluation criteria we identified for the *TARGET* taxonomy. Therefore, these metrics give an incomplete overview of CRX evaluation. Since we believe the identified metrics can still provide interesting insights for CRX designers, they are made publicly available in the GitHub repository referenced above. In addition, we would like to refer to the work of Chowdhury et al. [15] who conducted a literature review on cybersecurity exercises for critical infrastructure and extracted several evaluation metrics from these works.

(3) **Definition of Procedure:** Once suitable metrics have been selected, the modalities of the data collection for the evaluation need to be specified. This includes, for example, the preparation of questionnaires and the technical implementation of monitoring components in the CRX infrastructure, e.g., modules for automated assessment of the trainees' performance in the CRX. Furthermore, ethical considerations regarding a planned user study should be clarified in this step of the process.

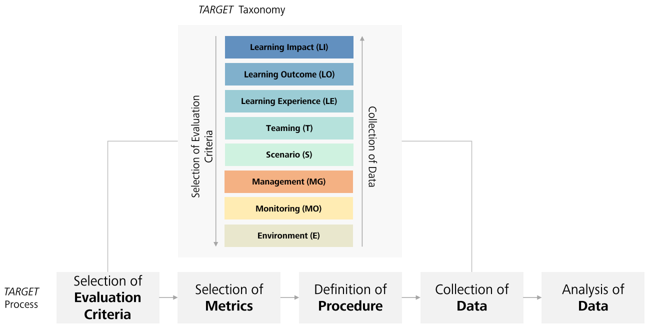


Figure 5: The five-step *TARGET* process: The process consists of the sequential steps Selection of Evaluation Criteria, Selection of Metrics, Definition of Procedure, Collection of Data, and Analysis of Data. The selection of evaluation criteria is performed in the reverse order of their assessment.

(4) **Collection of Data:** As shown in Figure 5, data is collected in the reverse order of the previous step *Selection of evaluation criteria*. Evaluation criteria in the dimensions *Environment*, *Monitoring*, *Management*, and *Scenario* refer to characteristics and functionalities of the CRX that can be measured without trainees' interaction with the CRX. Hence, this data can be collected before conducting the CRX itself. Data for the assessment of evaluation criteria in the areas of *Teaming*, *Learning Experience*, *Learning Outcome*, and *Learning Impact* requires conducting a user study to assess the trainees' behavior in the CRX. Thereby, data is collected both during and after the training.

(5) **Analysis of Results:** In the final step, the results of the evaluation are analyzed and discussed. At this point, it should be defined in which areas the CRX is performing as desired and in which areas there is a need for improvement to derive measures for implementing these improvements.

7 ICEBERG: A CRX FOR INCIDENT RESPONSE TRAINING

To demonstrate the benefit of the *TARGET* framework, we apply it to evaluate Iceberg CRX, which we developed to provide students in undergraduate and graduate cybersecurity courses a practical insight into cybersecurity. As a basis for understanding its evaluation, the following section briefly outlines the concept and implementation of Iceberg CRX. For more comprehensive explanations and detailed documentation of Iceberg CRX, we want to refer to the GitHub project[2] in which its source code is made publicly available.

7.1 Concept

In Iceberg CRX, trainees gain hands-on experience in incident response (IR), i.e., the ability to detect and handle cybersecurity incidents in an effective and systematic way [17]. IR is usually embedded in the processes of a security operations center (SOC), an organizational unit that oversees security operations to improve the organization's overall security posture [75]. The central technical

[2] https://github.com/TARGETframework

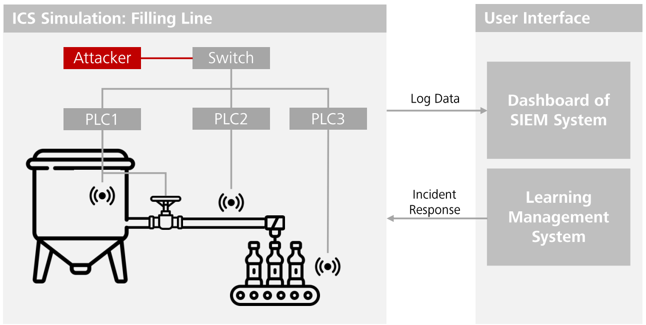
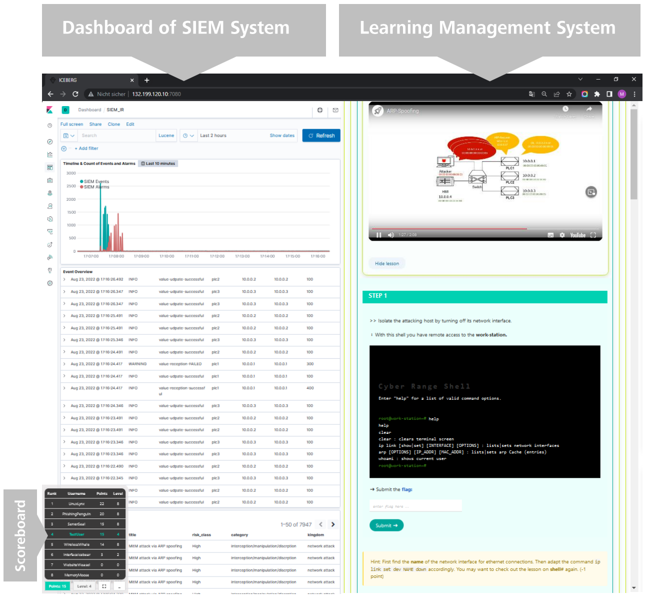


Figure 6: Schematic illustration of Iceberg CRX: The ICS simulation consists of a tank equipped with a motoric valve, a pipe, and a bottling station, each controlled by one PLC. The PLCs are interconnected over a switch, as is the simulated attacker. The simulation produces log data that is visible to the trainee over the dashboard of the SIEM system. Through the Learning Management System, the trainee can interact with the network of the ICS simulation.

platform of a SOC is a Security Information and Event Management (SIEM), providing an overarching solution to support the SOC's processes, which includes incident detection – one essential part of IR.

The scenario of Iceberg CRX is located in an industrial setting. Industrial control systems (ICS) integrate elements from OT and IT. Connecting former isolated industrial assets via a network, possibly even the Internet, leads to new attack vectors that are relevant to address through cybersecurity training. In Iceberg CRX, trainees take on the role of incident responders in a SOC, overseeing the security of an industrial filling line composed of a tank storing liquid to be filled into bottles via a pipe (rf. Figure 6). Three sensors, each controlled by a programmable logic controller (PLC), measure the liquid level inside the tank (PLC1), the flow rate of the pipe (PLC2), and the liquid level inside the bottle (PLC3). PLC1 additionally controls a motoric valve attached to the tank. The three PLCs are connected via a switch. An intrusion detection system (IDS) is running on each host (i.e., PLCs) to detect anomalies in the network traffic. The PLCs produce logs of operational events of the physical process (e.g., whenever a sensor captures data) as well as logs about monitoring activities of the IDS. This log data is fed into the SIEM system with which the trainees interact. In our scenario, a simulated attacker has gained access to the ICS's network and interferes with the filling process by performing a man-in-the-middle (MITM) attack between PLC1 and PLC3. By disrupting the communication between those PLCs, PLC3 is operating normally, yet PLC1 does not receive the sensor value of PLC3 to determine when to close the tank's valve resulting in an overflow of the bottles.

Trainees participate in Iceberg CRX over its web-based UI (rf. Figure 7), consisting of a dashboard that is part of the SIEM system and monitors the filling line and a Learning Management System (LMS) guiding the trainees through the training. The LMS entails both practical tasks and theoretical lessons. The lessons are text- or video-based units in which the trainee is provided background knowledge on IR and the ICS network. In eight sequential tasks,



Event Overview

```
> Aug 23, 2022 @ 17:16:26.492  INFO     value-udpate-successful  plc2  10.0.0.2  10.0.0.2  100
> Aug 23, 2022 @ 17:16:26.347  INFO     value-udpate-successful  plc3  10.0.0.3  10.0.0.3  100
> Aug 23, 2022 @ 17:16:26.347  INFO     value-udpate-successful  plc3  10.0.0.3  10.0.0.3  100
> Aug 23, 2022 @ 17:16:25.491  INFO     value-udpate-successful  plc2  10.0.0.2  10.0.0.2  100
> Aug 23, 2022 @ 17:16:25.491  INFO     value-udpate-successful  plc2  10.0.0.2  10.0.0.2  100
> Aug 23, 2022 @ 17:16:25.346  INFO     value-udpate-successful  plc3  10.0.0.3  10.0.0.3  100
> Aug 23, 2022 @ 17:16:24.491  INFO     value-udpate-successful  plc2  10.0.0.2  10.0.0.2  100
> Aug 23, 2022 @ 17:16:24.417  WARNING  value-reception-FAILED   plc1  10.0.0.1  10.0.0.1  300
> Aug 23, 2022 @ 17:16:24.417  INFO     value-udpate-successful  plc1  10.0.0.1  10.0.0.1  100
> Aug 23, 2022 @ 17:16:24.417  INFO     value-reception-successf plc1  10.0.0.1  10.0.0.1  400
                                        ul
> Aug 23, 2022 @ 17:16:24.346  INFO     value-udpate-successful  plc3  10.0.0.3  10.0.0.3  100
> Aug 23, 2022 @ 17:16:23.491  INFO     value-udpate-successful  plc2  10.0.0.2  10.0.0.2  100
> Aug 23, 2022 @ 17:16:23.491  INFO     value-udpate-successful  plc2  10.0.0.2  10.0.0.2  100
> Aug 23, 2022 @ 17:16:23.346  INFO     value-udpate-successful  plc3  10.0.0.3  10.0.0.3  100
> Aug 23, 2022 @ 17:16:23.346  INFO     value-udpate-successful  plc3  10.0.0.3  10.0.0.3  100
> Aug 23, 2022 @ 17:16:22.490  INFO     value-udpate-successful  plc2  10.0.0.2  10.0.0.2  100
> Aug 23, 2022 @ 17:16:22.345  INFO     value-udpate-successful  plc3  10.0.0.3  10.0.0.3  100
                               INFO     value-udpate-successful  plc2  10.0.0.2  10.0.0.2  100
                               INFO     value-udpate-successful  plc2  10.0.0.2  10.0.0.2  100
```

title	risk_class	category	kingdom
MitM attack via ARP spoofing	High	interception/manipulation/discrption	network attack
MitM attack via ARP spoofing	High	interception/manipulation/discrption	network attack
MitM attack via ARP spoofing	High	interception/manipulation/discrption	network attack
MitM attack via ARP spoofing	High	interception/manipulation/discrption	network attack

Rank	Username	Points	Level
1	LinuxLynx	22	8
2	PhishingPenguin	20	8
3	ServerSeal	15	8
4	TestUser	15	4
5	WirelessWhale	14	8
6	InterfaceIcebear	3	2
7	WebsiteWeasel	0	0
8	MemoryMoose	0	0

STEP 1

>> Isolate the attacking host by turning off its network interface.

ℹ With this shell you have remote access to the **work-station.**

```
Cyber Range Shell
Enter "help" for a list of valid command options.

root@work-station~# help
help
clear
clear : clears terminal screen
ip link [show|set] [INTERFACE] [OPTIONS] : lists|sets network interfaces
arp [OPTIONS] [IP_ADDR] [MAC_ADDR] : lists|sets arp Cache (entries)
whoami : shows current user
root@work-station~#
```

→ Submit the **flag:**

enter flag here ..

Submit →

Hint: First find the **name** of the network interface for ethernet connections. Then adapt the command `ip link set dev NAME down` accordingly. You may want to check out the lesson on **shell#** again. (-1 point)

Figure 7: UI of Iceberg CRX: The UI is composed of two parts, the dashboard of the SIEM system displaying incoming security alarms and events and the Learning Management System, guiding the trainee through the CRX with a sequence of lessons and tasks.

the trainee investigates log data in the SIEM system to detect the MITM attack and accesses the network over a simulated command line to eliminate the attacker and restore the network. By engaging in the tasks, the trainees acquire hands-on IR skills. The CRX includes several gamification elements to raise the trainees' learning motivation. For each task the trainees solve, they collect points and achieve a new level unlocking the next lesson and task. To further gamify the training, a scoreboard in the LMS displays the points and levels of all trainees participating in the current run of the CRX, enabling a trainee to compare their performance to those of the other trainees.

7.2 Implementation

To make the implementation of Iceberg CRX both time- and cost-efficient, we decided to reuse existing CRX components where possible instead of developing the CRX from scratch. For this reason, we designed Iceberg CRX building upon a CRX by Vielberth et al. [76], which uses the integration of an ICS simulation in a SOC to teach trainees how to configure a SIEM system. We utilize this environment for a novel training focus (IR instead of SIEM configuration).

The overall architecture relies on a microservice infrastructure implemented with several Docker[3] containers and is deployed on one VM for each trainee or team of trainees participating. The simulation of the filling line is implemented with MiniCPS[4], an extension of the python-based network simulation tool Mininet[5] that enables

[3] https://www.docker.com/
[4] https://github.com/scy-phy/minicps
[5] http://www.mininet.org

the simulation of ICSs. The MITM attack is realized with Ettercap[6], a tool for penetration testing. The SIEM functions are implemented with Dsiem[7], an open-source SIEM system that builds upon the ELK stack[8] and provides a Kibana-based dashboard for visualizing incoming events. The user interface is implemented with the JavaScript Framework VueJS[9] and consists of the SIEM dashboard and the LMS, displaying learning material and the trainees' tasks. For a more detailed description of this architecture, we want to refer to the work of Vielberth et al. [76].

For Iceberg CRX, we extended the existing simulation in several ways. Firstly, we enhanced the ICS system with incident detection functionalities in the form of an IDS running on each PLC, monitoring the network traffic. The IDS is implemented with the python-based network tool scapy[10]. Secondly, the simulation of the MITM attack was improved to automatically execute the attack based on the trainee's progress in the CRX. The new training focus also required the development of entirely new training materials for the LMS. For example, we developed a new type of task in which the trainees use a simulated terminal (rf. Figure 7) to interact with the ICS simulation.

8 INSTANTIATION OF THE *TARGET* FRAMEWORK: EVALUATION OF ICEBERG CRX

In this section, we instantiate the *TARGET* framework to evaluate Iceberg CRX, which is the final step of this work's method introduced in Section 4. Besides demonstrating the framework's applicability, this shall also serve as a guiding example for CRX designers.

8.1 Selection of Evaluation Criteria

The goal of the evaluation is to investigate whether Iceberg CRX can positively contribute to academic cybersecurity education and, thus, whether the CRX should be permanently included in two cybersecurity courses at our university. In this regard, we first selected relevant evaluation criteria from the *TARGET* taxonomy and further specified each criterion for Iceberg CRX. The selected criteria are listed in Table 2. Due to the limited scope of this paper, we could only include a few criteria in the evaluation. As we believe the *Learning* dimensions to have shown the most significant results, we focused on the description of criteria in these dimensions. We are aware that this results in the other dimensions only being evaluated to a small extent. Our set of criteria should still be adequate to demonstrate the framework's application, which is the purpose of this instantiation.

8.2 Selection of Metrics

8.2.1 Learning Impact. The trainees' **Willingness for continuous education and training (LI.1)** was assessed by (1) collecting qualitative feedback by asking the trainees to give their impression of Iceberg CRX either orally or through a free-text form and by (2) assessing how many of the trainees signed up to a mailing list to participate in future CRXs.

8.2.2 Learning Outcome. To assess the fulfillment of the four criteria **Non-security related technical knowledge (LO.8)**, **Offensive security knowledge (LO.1)**, **Defensive security knowledge (LO.3)** and **Defensive security skills (LO.4)**, we measured if participation in Iceberg CRX leads to an increase in skills and knowledge. Because no comparable IR training was previously part of the courses, a control group design was not feasible. Instead, a one-group pretest/posttest design was used to assess the trainees' skills and knowledge before and after training. The metric resulting from this design is the difference (in percent) of mean correctly answered questions in the pretest and the posttest for each criterion, analyzed using a paired t-test. To assess the trainees' skills and knowledge, we developed a scale for each criterion consisting of a set of items in the form of multiple-choice questions. The method of using a multiple-item scale to measure a construct is seen to provide increased reliability and construct validity compared to a single-item method (i.e., one item respectively multiple-choice question to measure one criterion) [23].

8.2.3 Learning Experience. **Engagement of trainees (LE.2)** was assessed using the ARCS model by Keller [36], which utilizes the four categories Attention, Relevance, Confidence, and Satisfaction to measure learning motivation. Attention refers to catching and keeping the learner's interest throughout a learning intervention without over- or under-stimulating the learner. Relevance describes if the learning content relates to the learner's future activities and goals. The Confidence condition is met when a learner has the impression of succeeding with a reasonable effort. The Satisfaction condition is met when learners feel content about their achievements. The subjective perception of whether a learning exercise led to an increase in knowledge is seen as another factor that contributes to the trainee's intrinsic motivation to learn [24]. Thus, we complemented the ARCS model with the category Metacognition. Following the multiple-item method described above, we developed a scale to measure each condition with two items in the form of statements. The trainees were asked to rate their level of agreement with each of these statements on a Likert scale from 1 (completely disagree) to 5 (completely agree). To measure if the duration of the training is appropriate (**Adequacy of duration (LE.4)**), the time it takes trainees to complete the training was recorded. Finally, the difficulty of the Iceberg CRX (**Adequacy of difficulty (LE.3)**) was captured by the scores that trainees achieved during the training and the trainee's subjective perception of the CRX's difficulty, assessed through a feedback statement in the same manner as **LE.2**.

8.2.4 Scenario, Teaming, Management, Monitoring, and Environment. The evaluation criteria of the remaining dimensions primarily relate to functional properties, respectively, requirements of the CRX. Whether or not these criteria are achieved depends on the extent to which the particular functional properties were considered in the design and implementation of the CRX. These criteria, e.g., if the CRX enables automated monitoring of the trainees' actions, are hard to measure quantitatively but require a functional description of how the respective criterion is realized through the design and implementation of the CRX. This description serves as a basis for

[6] https://www.ettercap-project.org/
[7] https://github.com/defenxor/dsiem
[8] https://www.elastic.co
[9] https://vuejs.org/
[10] https://scapy.net/

Table 2: Criteria Definition for Iceberg CRX.

No.	Evaluation Criterion	Description
LI.1	**Willingness for continuous education and training**	Iceberg CRX raises the trainees' interest in cybersecurity, more specifically IR, motivating them to engage more with the topic and consider a career path in cybersecurity in the future.
LO.1	**Offensive security knowledge**	The trainees gain background knowledge on MITM attacks on the network layer.
LO.3	**Defense security knowledge**	The trainees gain background knowledge on the IR life cycle.
LO.4	**Defensive security skills**	The trainees learn to detect certain attacks by investigating log data within a SIEM system. Furthermore, they learn to use command line instructions to eliminate a host from the network and restore ARP tables in the network.
LO.8	**Non-security related technical knowledge**	The trainees gain knowledge of basic ICS-related terms (e.g., PLC) and an exemplary network structure of an ICS.
LE.2	**Engagement of trainees**	Iceberg CRX is capable of motivating trainees to engage with the training.
LE.3	**Adequacy of difficulty**	The difficulty level of the CRX is appropriate for the intended audience (students in a cybersecurity course) so that learners do not feel overwhelmed by the CRX but can successfully complete the majority of the tasks.
LE.4	**Adequacy of duration**	Iceberg CRX can be completed within one course session (90 minutes).
T.2	**Dynamic teaming roles configuration**	Trainees are able to take on the role of the Blue Team individually or, alternatively, in small teams.
S.1	**Relevance**	The ICS and SOC processes that form the CRX scenario are realistic.
S.5	**Fidelity of tools**	Trainees are using tools they would use in the same or similar way in real-life incident response.
S.7	**Adaptability of difficulty**	Trainees can adjust the difficulty of the training if they have trouble finding the solution to a particular task.
S.8	**Immediate feedback for trainees**	The trainees get immediate feedback for their actions during the CRX to ensure the flow of training.
MG.6	**Automation of attacks**	The attacks are automatically triggered and do not require the trainer's intervention.
MO.2	**Automation of trainee monitoring**	The trainees' training progress can be automatically monitored to identify if trainees have problems progressing with the training.
E.4	**Reliability**	Iceberg CRX's infrastructure runs stable throughout the training.
E.17	**Privacy**	The trainees' personal data is processed in a privacy-protecting manner.

deriving the degree to which the respective goal has been met in the *TARGET* process step *Analysis of Data*. This applies to the criteria **Dynamic teaming roles configuration (T.2), Relevance (S.1), Fidelity of tools (S.5), Adaptability of difficulty (S.7), Immediate feedback for trainees (S.8), Automation of attacks (MG.6), Automation of trainee monitoring (MO.2)**, and **Privacy (E.17)**. As the criterion **Reliability (E.4)** depends on not only the implementation of the CRX but also the technical infrastructure the CRX is deployed on, reliability is assessed through the number of VMs that encounter technical failures during a dry run of the training.

8.3 Definition of Procedure

The evaluation procedure is illustrated in Figure 8. Prior to the user study, the functional description of the CRX with regard to the individual evaluation criteria was created (**T.2, S.1, S.5, S.7, S.8, MG.6, MO.2, E.17**). The infrastructure was deployed on 15 VMs for the user study, enabling 15 individual trainees or teams of trainees participating in the CRX at the same time. For the dry run assessing the infrastructure's reliability (**E.4**), we deployed Iceberg CRX on the VMs and let the simulations run for three hours. To evaluate *Learning Outcome* the trainees' skills and knowledge (**L0.1, LO.3, LO.4, LO.8**) were assessed directly before and after the training (pre- and posttest). To assess the trainees' learning experience, they were asked to rate the overall 11 items described in Subsection 8.2.3 in the form of feedback statements (**LE.2, LE.3**) and describe their general impression of the training's value in a free-text form (**LI.1**). Finally, the trainees were informed verbally to email us if they wanted to be notified of upcoming CRX events. The planned user study was designed in accordance to the guidelines of the ethics committee of our university and did not raise any ethical concerns. Thus, no particular actions had to be taken in this regard.

8.4 Collection of Data

The user study was conducted with 50 participants recruited from undergraduate and graduate cybersecurity courses at a German university. Of the 50 participants, 38 identified as male and 12 as female, 35 were undergraduate, and 15 were graduate students. The user study was conducted in six runs, with three runs with students recruited from the undergraduate course in December 2021 and three runs with students recruited from the graduate course in May

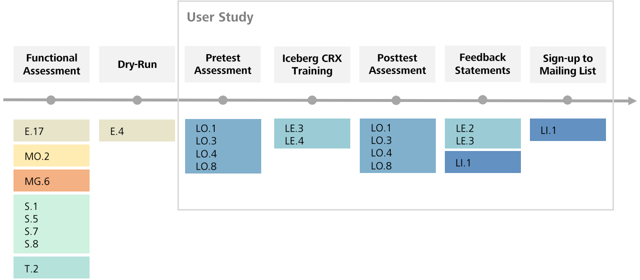


Figure 8: Procedure of Iceberg CRX's evaluation: The evaluation includes a functional assessment, a dry-run and a user study. The user study consists of the five sub steps Pretest assessment, the CRX Training, Posttest assessment, Assessment of feedback statements, and Sign up to the mailing list.

2022. Since no in-class training was possible in December 2021 due to restrictions during the COVID-19 pandemic, participants attended remotely from home. To ensure comparable conditions for all training runs, the training runs in May were also conducted remotely. Four of the 50 participants did not complete the feedback survey concerning the **Engagement of trainees (LE.2)**, resulting in a data set of n=46 for this criterion. Due to technical problems, times were only correctly recorded for n=40 participants. To provide high reproducibility and reusability, our data set and the evaluation questionnaires are made publicly available[11].

8.5 Analysis of Results

Since describing the evaluation results of each criterion in detail would exceed the scope of this paper, we limit the detailed analysis to the categories for which the evaluation results were particularly noteworthy. For the remaining criteria, we would like to refer to Table 4, which summarizes the evaluation results, and to our GitHub repository, which provides detailed descriptions of all evaluation results.

LI.1 Willingness for continuous education and training: Sample feedback we received from the trainees was "The training was very interesting and gave me some interesting insights how this tool works and how attacks do look like in real life. Please more of this!", "The Cyber Range was great, motivated to go deeper into IT-Sec" and "A very cool interface with a lot of useful and probably realistic tools. I really liked the playful idea of teaching and having like a competition against the attacker. A great training that definitely grew my interest in those areas". None of the participants' feedback indicated that they found the training uninteresting or irrelevant. 21 out of the 50 students (42%) participating in the user study actively signed up to our mailing list signalizing they want to participate in future CRXs. These results indicate that Iceberg CRX generally has the potential to raise students' interest in cybersecurity and encourage students to learn more about the topic in the future. However, long-term impacts, such as students enrolling in cybersecurity master programs or starting a cybersecurity career in the industry, need to be measured by further studies.

LO.1 Offensive security knowledge: Offensive security knowledge significantly increased by 42% from pretest ($M = .61, SD = .25$) to posttest ($M = .87, SD = .20$), $t(49) = -6, 38, p < 0.001$.

LO.3 Defensive security knowledge: Defensive security knowledge significantly increased by 19% from pretest ($M = .69, SD = .26$) to posttest, ($M = .82, SD = .20$), $t(49) = -4.39, p < 0.001$.

LO.4 Defensive security skills: With a significant increase 70% from pretest ($M = .56, SD = .28$) to posttest ($M = .95, SD = .17$), the highest impact of Iceberg CRX regarding a change in the participants' skills and knowledge was observed for this criterion, $t(49) = -9.31, p < 0.001$. This specifically highlights Iceberg CRX's potential to convey hands-on cybersecurity skills.

LO.8 Non security-related technical knowledge: Non security-related knowledge significantly increased by 29% from pretest ($M = .55, SD = .30$) to posttest, ($M = .71, SD = .24$), $t(49) = -3.63, p < 0.001$.

Across all four criteria assessed within the pre-/posttest design, the results of the t-test showed a significantly better result of the posttest ($M = .83, SD = .13$) compared to the pretest ($M = .62, SD = .18$), increasing by 34%, $t(49) = -10.06, p < .001$. This indicates that participation in the cyber range training led to an overall increase in skills and knowledge among the participants. Figure 9 shows the results of the overall learning outcome and the results for each criterion.

LE.2 Engagement of trainees: With an overall mean of 4.09 ($SD = .84, Mdn = 4$), the participants' intrinsic motivation for the training can be considered high. The results for each category are depicted in Table 3 and Figure 10.

LE.3 Adequacy of difficulty: On average, the participants scored 15.9 out of 24 points ($SD = 6.17, Mdn = 17$), indicating that participants were, in general, able to solve a majority of the tasks. The participants' mean rating of the statement if they found the tasks of the CRX overwhelming was 2.43 ($SD = .94$) and a median of 2 ("disagree"). The relatively high standard deviation, however, suggests that some participants were overwhelmed by the difficulty of the training. This was also reflected in the feedback from the participants, who felt put off by the complexity of the SIEM dashboard at the beginning of the training. It can be concluded from this that the difficulty level of Iceberg CRX, in part, is too high for the target group.

S.5 Fidelity of tools: The SIEM tool Dsiem is part of the CRX by Vielberth et al. [76] which we utilized as a base for Iceberg CRX. To date, Dsiem is no more actively developed, and technical support is limited. For this reason, Dsiem is now highly unlikely to be used in practice, which is why the fidelity of the tool is limited.

MO.2 Automation of trainee monitoring: Participants' scores are stored in a document-based database that provides a user interface for displaying and filtering the documents. This user interface can be used to monitor trainees' progress

[11] https://github.com/TARGETframework/IcebergEvaluation

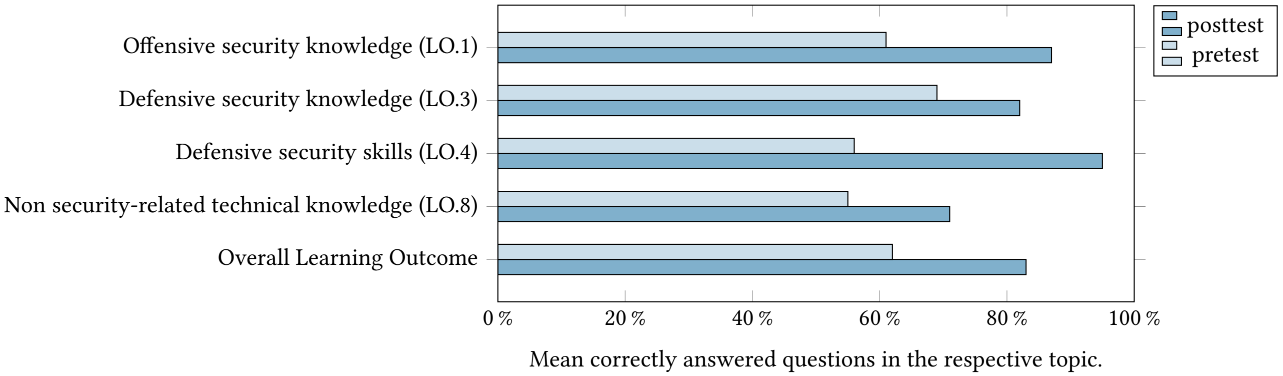


Figure 9: Evaluation results in the dimension Learning Outcome: The participants' mean increase in skills and knowledge was over 19% in every assessed topic.

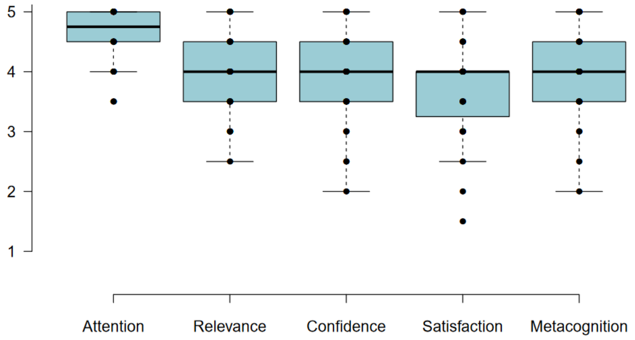


Figure 10: Evaluation results in the dimension learning outcome: Each assessed aspect showed a median of 4 or 5. The highest values were observed in the category "Attention", the lowest values were assessed in the "Satisfaction" category, which also has the highest level of scattering.

Table 3: Results of the feedback statements assessing the participants' engagement.

Condition	Mean	Median	Standard Deviation
Attention	4.64	5	.60
Relevance	4.00	4	.75
Confidence	3.99	4	.80
Satisfaction	3.75	4	.96
Metacognition	4.07	4	.77

during training, for example, to display how many trainees have already reached a certain level in training. This type of monitoring was sufficient for the training sessions in the user study with a maximum of 15 participants. However, to be able to monitor a larger number of trainees, a better-automated monitoring system would be necessary.

E.4 **Reliability:** During the dry run, the composition of the Docker container failed at two machines. This could be fixed, however, by simply restarting the infrastructure. Once the Docker containers were running on each machine, the Iceberg CRX instances were functioning correctly throughout the dry run. Hence, the system's reliability was considered sufficient, confirmed in the user study, as none of the participants faced any technical difficulties related to the infrastructure. The only problem we faced was monitoring the trainees, as the time recording was incomplete for ten participants.

To summarize, the evaluation validated the strengths of Iceberg CRX and identified current weaknesses that point the way to future improvement. The positive evaluation results in the three *Learning* dimensions indicate that Iceberg CRX can positively contribute to improving academic cybersecurity education. However, in order to make the training accessible to a larger number of students, some technical adjustments have to be made, especially regarding the monitoring of trainees. In addition, the SIEM system, which is currently under use, should be replaced with a more modern tool to ensure higher a level of fidelity.

8.6 *TARGET* Framework's Potential in the Light of the Instantiation

Finally, we summarize the instantiation of the *TARGET* framework by highlighting how the evaluation of Iceberg CRX has benefited from its alignment with the framework. From our experience, the utility of the *TARGET* framework was particularly evident in two aspects. Having the taxonomy as a knowledge base for selecting evaluation criteria relevant to Iceberg CRX made us include several criteria in the evaluation that we were initially not aware of. Designing the CRX, we had quite a clear idea of what trainees should learn (*Learning Outcome*) and that training should be motivating and engaging (*Learning Experience*). However, we initially gave little thought to what underlying technical aspects the CRX needs to provide to enable these Learning-related objectives. For example, the taxonomy made us realize that for Iceberg CRX, the monitoring of trainees (**MO.2**) is a critical factor. Unless monitoring is at least partially automated, only a small number of trainees can be overseen at once. However, the intended use of Iceberg CRX in a university course requires that many students can participate in the

Table 4: Overview of evaluation results: Evaluation criteria that are ranked as sufficiently fulfilled (●), partly fulfilled (◑), or not fulfilled (○)

ID	Evaluation Criterion	Fulfilled	Implications
LI.1	**Willingness for continuous education and training**	◑	First results are promising, yet assessing Iceberg CRX's long-term learning impact requires further evaluation.
LO.1	**Offensive security knowledge**	●	-
LO.3	**Defense security knowledge**	●	-
LO.4	**Defensive security skills**	●	-
LO.7	**Non-security related technical knowledge**	●	-
LE.2	**Engagement of trainees**	●	-
LE.3	**Adequacy of difficulty**	◑	The introduction to the training should be made easier for the trainees, e.g. by a more detailed introduction to the SIEM dashboard and the command line interface.
LE.4	**Adequacy of duration**	●	-
T.2	**Dynamic teaming roles configuration**	◑	Even though the design of Iceberg CRX theoretically allows trainees to participate not only individually but also in small groups, further evaluation is still needed to test whether this configuration leads to comparably good results in the *Learning* dimensions.
S.1	**Relevance**	●	-
S.5	**Fidelity of tools**	○	The currently used SIEM system Dsiem should be replaced by a more advanced and up-to-date SIEM solution.
S.7	**Adaptability of difficulty**	●	-
S.8	**Immediate feedback for trainees**	●	-
MG.6	**Automation of attacks**	●	-
MO.2	**Automation of trainee monitoring**	◑	In order to facilitate monitoring and allow a higher number of trainees participating at the same time, automated monitoring of participant actions and progress should be implemented.
E.4	**Reliability**	◑	Although the reliability of the infrastructure was sufficient for the number of participants studied, it is necessary to investigate why the set up of Docker containers partially failed. Additionally, the time recording module should be fixed.
E.17	**Privacy**	●	-

training simultaneously. While the *TARGET* taxonomy contributed to more multifaceted evaluation results, the top-down approach of the *TARGET* process provided us with an efficient evaluation design. When previously evaluating other CRXs we designed, we often did not have a clear idea of which aspects to evaluate from the beginning. This led us to capture far more metrics than necessary, resulting in an unnecessarily long and complex evaluation for the trainees and an inefficient analysis of the collected data. Determining metrics based on the selected evaluation criteria helped us to only assess what was informative.

9 LIMITATIONS AND FUTURE WORK

Our work is subject to some limitations we want to acknowledge in this section and outline how they can be addressed in future work. The taxonomy we present as part of the *TARGET* framework is clearly only as good as the underlying literature review. As we limited our search to four academic online libraries, potentially relevant papers from other libraries or CRX concepts from commercial vendors were not considered. Due to this potential selection bias, the completeness of the taxonomy may be questioned. The papers we identified in our search nevertheless represent a wide variety of different CRX concepts and, in our estimation, incorporate the most relevant work in the field. While we believe that the extracted evaluation criteria can provide CRX designers with a solid overview of what aspects might be generally relevant in a CRX evaluation, this set of criteria is not exhaustive and can only represent a current snapshot of CRX research. The taxonomy should therefore be seen as a dynamic construct that can be adapted and extended in the future. The second limitation lies in the evaluation process itself. Which criteria are considered relevant and therefore included in the evaluation is subject to the stakeholders' decision. Admittedly, this can lead to some criteria, though perhaps interesting or relevant, yet not the stakeholder's main focus, not being assessed. Nevertheless, we think that only through the flexibility the selection of criteria provides the framework can be used for CRX with different foci. The subject of future research could be to develop a minimal set of criteria that should be included in any CRX evaluation showing which criteria should be prioritized.

Regarding the operationalization of the framework, it can be criticized that in its current form, it provides little guidance for CRX designers on selecting metrics and developing measurement instruments to assess the criteria. The instantiation of the framework described in Section 8 intends to demonstrate how the TARGET framework can be applied to a particular CRX. Due to the brevity of the description of this instantiation, the process for selecting appropriate metrics and developing the scales used for the trainees' self-assessment has only been described in brief. As a result, the described operationalization misses, for example, the validation of the internal consistency of the individual items of the developed scales used for the self-assessment of the trainees. Extending the framework with a mapping of possible metrics to the taxonomy criteria will be the subject of future research. The operationalization of the framework by CRX designers in research and practice will provide essential insights in this regard.

Finally, to date, the *TARGET* process only offers the possibility for summative evaluation of a CRX, i.e., the evaluation of a ready-to-use CRX. Including the taxonomy in a process for formative evaluation in the future would have the potential to improve a CRX already during its design and development phase. For this reason, the *TARGET* process should be extended in this regard.

10 CONCLUSION

In this work, we present a framework that aims to structure and facilitate CRX evaluation. For this matter, we developed a literature-based taxonomy of evaluation criteria for CRXs and embedded the taxonomy into a five-step evaluation process. The taxonomy serves as a knowledge base for CRX designers guiding them on which aspects to consider in evaluating a CRX. To demonstrate the framework's applicability, we instantiated it to evaluate Iceberg CRX, a CRX to provide students in cybersecurity courses at our university with hands-on practice in responding to cybersecurity incidents. Applying our framework, we highlighted how structured evaluation can help identify the strengths and weaknesses of Iceberg CRX. While Iceberg CRX is very suitable for conveying students with incident response skills and knowledge and raising their interest, the underlying technical infrastructure can be improved in some regards. To conclude, we are certain that *TARGET* framework contributes to realizing the full potential of CRXs. As such, our research represents one advance in the quest to improve cybersecurity training and provides a base for other researchers to build upon.

ACKNOWLEDGMENTS

We want to thank Richard Schuster for his great work on the Iceberg prototype, the many fruitful discussions and his valuable comments on this manuscript. Furthermore, we want to thank Helen Petrie for looking at our research from a different perspective and providing us with insightful and constructive feedback.

This work was performed under the INSIST project, which is supported under contract by the Bavarian Ministry of Economic Affairs, Regional Development and Energy (DIK0338/01).

REFERENCES

[1] Jonas Almroth and Tommy Gustafsson. 2020. CRATE Exercise Control – A cyber defense exercise management and support tool. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 37–45. https://doi.org/10.1109/EuroSPW51379.2020.00014

[2] Mauro Andreolini, Vincenzo Giuseppe Colacino, Michele Colajanni, and Mirco Marchetti. 2020. A Framework for the Evaluation of Trainee Performance in Cyber Range Exercises. *Mobile Networks and Applications* 25, 1 (2020), 236–247. https://doi.org/10.1007/s11036-019-01442-0

[3] Markos Antonopoulos, Giorgos Drainakis, Eletherios Ouzounoglou, Giorgos Papavassiliou, and Angelos Amditis. 2022. Design and proof of concept of a prediction engine for decision support during cyber range attack simulations in the maritime domain. In *2022 IEEE International Conference on Cyber Security and Resilience (CSR)*. 305–310. https://doi.org/10.1109/CSR54599.2022.9850280

[4] Sobia Arshad, Masoom Alam, Saif Al-Kuwari, and Muhammad Haider Ali Khan. 2021. Attack Specification Language: Domain Specific Language for Dynamic Training in Cyber Range. In *2021 IEEE Global Engineering Education Conference (EDUCON)*. 873–879. https://doi.org/10.1109/EDUCON46332.2021.9454094

[5] Victor R Basili. 1994. Goal question metric paradigm. *Encyclopedia of software engineering* (1994), 528–532.

[6] Giorgio Bernardinetti, Stefano Iafrate, and Giuseppe Bianchi. 2021. Nautilus: A Tool For Automated Deployment And Sharing Of Cyber Range Scenarios. In *The 16th International Conference on Availability, Reliability and Security* (Vienna, Austria) *(ARES 2021)*. Association for Computing Machinery, New York, NY, USA, Article 146, 7 pages. https://doi.org/10.1145/3465481.3469182

[7] Razvan Beuran, Takuya Inoue, Yasuo Tan, and Yoichi Shinoda. 2019. Realistic Cybersecurity Training via Scenario Progression Management. In *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 67–76. https://doi.org/10.1109/EuroSPW.2019.00014

[8] Razvan Beuran, Dat Tang, Cuong Pham, Ken ichi Chinen, Yasuo Tan, and Yoichi Shinoda. 2018. Integrated framework for hands-on cybersecurity training: CyTrONE. *Computers & Security* 78 (2018), 43–59. https://doi.org/10.1016/j.cose.2018.06.001

[9] Razvan Beuran, Zhe Zhang, and Yasuo Tan. 2022. AWS EC2 Public Cloud Cyber Range Deployment. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 433–441. https://doi.org/10.1109/EuroSPW55150.2022.00051

[10] Chiara Braghin, Stelvio Cimato, Ernesto Damiani, Fulvio Frati, Elvinia Riccobene, and Sadegh Astaneh. 2020. Towards the Monitoring and Evaluation of Trainees' Activities in Cyber Ranges. In *Model-driven Simulation and Training Environments for Cybersecurity*, George Hatzivasilis and Sotiris Ioannidis (Eds.). Springer International Publishing, Cham, 79–91.

[11] Norbou Buchler, Prashanth Rajivan, Laura R. Marusich, Lewis Lightner, and Cleotilde Gonzalez. 2018. Sociometrics and observational assessment of teaming and leadership in a cyber security defense competition. *Computers & Security* 73 (2018), 114–136. https://doi.org/10.1016/j.cose.2017.10.013

[12] Francesco Caturano, Nicola d'Ambrosio, Gaetano Perrone, Luigi Previdente, and Simon Pietro Romano. 2022. ExploitWP2Docker: a Platform for Automating the Generation of Vulnerable WordPress Environments for Cyber Ranges. In *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. 1–7. https://doi.org/10.1109/ICECET55527.2022.9872859

[13] Francesco Caturano, Gaetano Perrone, and Simon Pietro Romano. 2020. Capturing flags in a dynamically deployed microservices-based heterogeneous environment. In *2020 Principles, Systems and Applications of IP Telecommunications (IPTComm)*. 1–7. https://doi.org/10.1109/IPTComm50535.2020.9261519

[14] Tianying Chen, Jessica Hammer, and Laura Dabbish. 2019. Self-Efficacy-Based Game Design to Encourage Security Behavior Online. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312935

[15] Nabin Chowdhury and Vasileios Gkioulos. 2021. Cyber security training for critical infrastructure protection: A literature review. *Computer Science Review* 40 (2021), 100361. https://doi.org/10.1016/j.cosrev.2021.100361

[16] Nabin Chowdhury, Sokratis Katsikas, and Vasileios Gkioulos. 2022. Modeling effective cybersecurity training frameworks: A delphi method-based study. *Computers & Security* 113 (2022), 102551. https://doi.org/10.1016/j.cose.2021.102551

[17] Paul Cichonski, Tom Millar, Tim Grance, and Karen Scarfone. 2012. *Computer security incident handling guide.* Technical Report 61. 1–147 pages.

[18] Lizzie Coles-Kemp, Rikke Bjerg Jensen, and Claude P. R. Heath. 2020. *Too Much Information: Questioning Security in a Post-Digital Society.* Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376214

[19] Michael Collins, Alefiya Hussain, and Stephen Schwab. 2022. Towards an Operations-Aware Experimentation Methodology. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 384–393. https://doi.org/10.1109/EuroSPW55150.2022.00046

[20] Melissa Dark and Jelena Mirkovic. 2015. Evaluation Theory and Practice Applied to Cybersecurity Education. *IEEE Security & Privacy* 13, 2 (2015), 75–80. https:

//doi.org/10.1109/MSP.2015.27

[21] Thibault Debatty and Wim Mees. 2019. Building a Cyber Range for training CyberDefense Situation Awareness. In *2019 International Conference on Military Communications and Information Systems (ICMCIS)*. 1–6. https://doi.org/10.1109/ICMCIS.2019.8842802

[22] Gary M. Deckard. 2018. Cybertropolis: breaking the paradigm of cyber-ranges and testbeds. In *2018 IEEE International Symposium on Technologies for Homeland Security (HST)*. 1–4. https://doi.org/10.1109/THS.2018.8574134

[23] R. F. . DeVellis. 1992. *Scale development: Theory and applications*. Sage Publications, Inc.

[24] Anastasia Efklides. 2011. Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational psychologist* 46, 1 (2011), 6–25.

[25] Demitrius Fenton, Terry Traylor, Guy Hokanson, and Jeremy Straub. 2019. Integrating Cyber Range Technologies and Certification Programs to Improve Cybersecurity Training Programs. In *The Challenges of the Digital Transformation in Education*, Michael E. Auer and Thrasyvoulos Tsiatsos (Eds.). Springer International Publishing, Cham, 632–643.

[26] Massimo Ficco and Francesco Palmieri. 2019. Leaf: An open-source cybersecurity training platform for realistic edge-IoT scenarios. *Journal of Systems Architecture* 97 (2019), 107–129. https://doi.org/10.1016/j.sysarc.2019.04.004

[27] Dorota Filipczuk, Charles Mason, and Stephen Snow. 2019. Using a Game to Explore Notions of Responsibility for Cyber Security in Organisations. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312846

[28] Sabrina Friedl, Magdalena Glas, Ludwig Englbrecht, Fabian Böhm, and Günther Pernul. 2022. ForCyRange: An Educational IoT Cyber Range for Live Digital Forensics. In *Information Security Education - Adapting to the Fourth Industrial Revolution*, Lynette Drevin, Natalia Miloslavskaya, Wai Sze Leung, and Suné von Solms (Eds.). Springer International Publishing, Cham, 77–91.

[29] Magdalena Glas, Manfred Vielberth, Tobias Reittinger, Fabian Böhm, and Günther Pernul. 2022. Visual Programming in Cyber Range Training to Improve Skill Development. In *Human Aspects of Information Security and Assurance*, Nathan Clarke and Steven Furnell (Eds.). Springer International Publishing, Cham, 3–13.

[30] Magdalena Granåsen and Dennis Andersson. 2016. Measuring team effectiveness in cyber-defense exercises: a cross-disciplinary case study. *Cognition, Technology & Work* 18, 1 (2016), 121–143.

[31] Tommy Gustafsson and Jonas Almroth. 2021. Cyber Range Automation Overview with a Case Study of CRATE. In *Secure IT Systems*, Mikael Asplund and Simin Nadjm-Tehrani (Eds.). Springer International Publishing, Cham, 192–209.

[32] George Hatzivasilis, Sotiris Ioannidis, Michail Smyrlis, George Spanoudakis, Fulvio Frati, Chiara Braghin, Ernesto Damiani, Hristo Koshutanski, George Tsakirakis, Torsten Hildebrandt, Ludger Goeke, Sebastian Pape, Oleg Blinder, Michael Vinov, George Leftheriotis, Martin Kunc, Fotis Oikonomou, Giovanni Magilo, Vito Petrarolo, Antonio Chieti, and Robert Bordianu. 2021. The THREAT-ARREST Cyber Range Platform. In *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*. 422–427. https://doi.org/10.1109/CSR51186.2021.9527963

[33] Olivier Jacq, Pablo Giménez Salazar, Kamban Parasuraman, Jarkko Kuusijärvi, Andriana Gkaniatsou, Evangelia Latsa, and Angelos Amditis. 2021. The Cyber-MAR Project: First Results and Perspectives on the Use of Hybrid Cyber Ranges for Port Cyber Risk Assessment. In *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*. 409–414. https://doi.org/10.1109/CSR51186.2021.9527968

[34] Mika Karjalainen and Tero Kokkonen. 2020. Comprehensive Cyber Arena; The Next Generation Cyber Range. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 11–16. https://doi.org/10.1109/EuroSPW51379.2020.00011

[35] Georgios Kavallieratos, Sokratis K Katsikas, and Vasileios Gkioulos. 2019. Towards a cyber-physical range. In *Proceedings of the 5th on Cyber-Physical System Security Workshop*. 25–34.

[36] John M Keller. 1987. Development and use of the ARCS model of instructional design. *Journal of instructional development* 10, 3 (1987), 2–10.

[37] Donald L. Kirkpatrick. 1967. *Training and Development Handbook*. American Society for Training and Development, New York: McGraw-Hill Book Company, Chapter Evaluation of Training.

[38] Stela Kucek and Maria Leitner. 2020. Training the Human-in-the-Loop in Industrial Cyber Ranges. In *Digital Transformation in Semiconductor Manufacturing*, Sophia Keil, Rainer Lasch, Fabian Lindner, and Jacob Lohmer (Eds.). Springer International Publishing, Cham, 107–118.

[39] Dennis Kundisch, Jan Muntermann, Anna Maria Oberländer, Daniel Rau, Maximilian Röglinger, Thorsten Schoormann, and Daniel Szopinski. 2021. An Update for Taxonomy Designers. *Business & Information Systems Engineering* (2021), 1–19.

[40] Claire La Fleur, Blaine Hoffman, C. Benjamin Gibson, and Norbou Buchler. 2021. Team performance in a series of regional and national US cybersecurity defense competitions: Generalizable effects of training and functional role specialization. *Computers & Security* 104 (2021), 102229. https://doi.org/10.1016/j.cose.2021.102229

[41] Xabier Larrucea and Alberto Molinuevo. 2020. An ICS Based Scenario Generator for Cyber Ranges. In *Systems, Software and Services Process Improvement*, Murat Yilmaz, Jörg Niemann, Paul Clarke, and Richard Messnarz (Eds.). Springer International Publishing, Cham, 543–554.

[42] Maria Leitner, Maximilian Frank, Wolfgang Hotwagner, Gregor Langner, Oliver Maurhart, Timea Pahi, Lenhard Reuter, Florian Skopik, Paul Smith, and Manuel Warum. 2020. AIT Cyber Range: Flexible Cyber Security Environment for Exercises, Training and Research. In *Proceedings of the European Interdisciplinary Cybersecurity Conference* (Rennes, France) *(EICC 2020)*. Association for Computing Machinery, New York, NY, USA, Article 2, 6 pages. https://doi.org/10.1145/3424954.3424959

[43] Stephen Lineberry. 2007. The human element: The weakest link in information security. *Journal of Accountancy* 204, 5 (2007), 44.

[44] Kaie Maennel, Rain Ottis, and Olaf Maennel. 2017. Improving and Measuring Learning Effectiveness at Cyber Defense Exercises. In *Secure IT Systems*, Helger Lipmaa, Aikaterini Mitrokotsa, and Raimundas Matulevičius (Eds.). Springer International Publishing, Cham, 123–138.

[45] Rasmi-Vlad Mahmoud, Egon Kidmose, Ahmet Turkmen, Olga Pilawka, and Jens Myrup Pedersen. 2021. DefAtt - Architecture of Virtual Cyber Labs for Research and Education. In *2021 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*. 1–7. https://doi.org/10.1109/CyberSA52016.2021.9478236

[46] Notis Mengidis, Maya Bozhilova, Cyril Ceresola, Consuelo Colabuono, Michael Cooke, Gregory Depaix, Angel Genchev, Georgi Koykov, Wim Mees, Matteo Merialdo, Antonis Voulgaridis, Theodora Tsikrika, Konstantinos Votis, and Stefanos Vrochidis. 2022. Leveraging Cyber Ranges for Prototyping, Certification and Training: The ECHO case. In *2022 IEEE International Conference on Cyber Security and Resilience (CSR)*. 299–304. https://doi.org/10.1109/CSR54599.2022.9850278

[47] Ryotaro Nakata and Akira Otsuka. 2021. CyExec*: A High-Performance Container-Based Cyber Range With Scenario Randomization. *IEEE Access* 9 (2021), 109095–109114. https://doi.org/10.1109/ACCESS.2021.3101245

[48] National Initiative for Cybersecurity Education (NICE). 2020. *The Cyber Range: A Guide*. Technical Report.

[49] Nikos Oikonomou, Notis Mengidis, Minas Spanopoulos-Karalexidis, Antonis Voulgaridis, Matteo Merialdo, Ivo Raisr, Kaarel Hanson, Paloma de La Vallee, Theodora Tsikrika, Stefanos Vrochidis, and Konstantinos Votis. 2021. ECHO Federated Cyber Range: Towards Next-Generation Scalable Cyber Ranges. In *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*. 403–408. https://doi.org/10.1109/CSR51186.2021.9527985

[50] Vittorio Orbinato. 2021. A next-generation platform for Cyber Range-as-a-Service. In *2021 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*. 314–318. https://doi.org/10.1109/ISSREW53611.2021.00094

[51] Radek Ošlejšek, Jan Vykopal, Karolína Burská, and Vít Rusňák. 2018. Evaluation of Cyber Defense Exercises Using Visual Analytics Process. In *2018 IEEE Frontiers in Education Conference (FIE)*. 1–9. https://doi.org/10.1109/FIE.2018.8659299

[52] Victor-Valeriu Patriciu and Adrian Constantin Furtuna. 2009. Guide for designing cyber security exercises. In *Proceedings of the 8th WSEAS International Conference on E-Activities and information security and privacy*. World Scientific and Engineering Academy and Society (WSEAS), 172–177.

[53] Adamantini Peratikou, Constantinos Louca, Stavros Shiaeles, and Stavros Stavrou. 2021. On Federated Cyber Range Network Interconnection. In *Selected Papers from the 12th International Networking Conference*, Bogdan Ghita and Stavros Shiaeles (Eds.). Springer International Publishing, Cham, 117–128.

[54] Cuong Pham, Dat Tang, Ken-ichi Chinen, and Razvan Beuran. 2016. CyRIS: A Cyber Range Instantiation System for Facilitating Security Training. In *Proceedings of the Seventh Symposium on Information and Communication Technology* (Ho Chi Minh City, Vietnam) *(SoICT '16)*. Association for Computing Machinery, New York, NY, USA, 251–258. https://doi.org/10.1145/3011077.3011087

[55] Mauno Pihelgas and Markus Kont. 2021. Frankenstack: Real-time Cyberattack Detection and Feedback System for Technical Cyber Exercises. In *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*. 396–402. https://doi.org/10.1109/CSR51186.2021.9527923

[56] Georgios Potamos, Adamantini Peratikou, and Stavros Stavrou. 2021. Towards a Maritime Cyber Range training environment. In *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*. 180–185. https://doi.org/10.1109/CSR51186.2021.9527904

[57] Maxime Puys, Pierre-Henri Thevenon, and Stéphane Mocanu. 2021. Hardware-In-The-Loop Labs for SCADA Cybersecurity Awareness and Training. In *The 16th International Conference on Availability, Reliability and Security* (Vienna, Austria) *(ARES 2021)*. Association for Computing Machinery, New York, NY, USA, Article 147, 10 pages. https://doi.org/10.1145/3465481.3469185

[58] Basili V. R. 1994. Goal Question Metric Paradigm. *Encyclopedia of Software Engineering* (1994), 528–532.

[59] Martina A. Rau, Vincent Aleven, Nikol Rummel, and Stacie Rohrbach. 2013. Why Interactive Learning Environments Can Have It All: Resolving Design Conflicts between Competing Goals. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for

Computing Machinery, New York, NY, USA, 109–118. https://doi.org/10.1145/2470654.2470670

[60] Elaine M. Raybourn, Michael Kunz, David Fritz, and Vince Urias. 2018. *A Zero-Entry Cyber Range Environment for Future Learning Ecosystems.* Springer International Publishing, Cham, 93–109. https://doi.org/10.1007/978-3-319-98935-8_5

[61] Filippo Rebecchi, Antonio Pastor, Alberto Mozo, Chiara Lombardo, Roberto Bruschi, Ilias Aliferis, Roberto Doriguzzi-Corin, Panagiotis Gouvas, Antonio Alvarez Romero, Anna Angelogianni, Ilias Politis, and Christos Xenakis. 2022. A Digital Twin for the 5G Era: the SPIDER Cyber Range. In *2022 IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM).* 567–572. https://doi.org/10.1109/WoWMoM54355.2022.00088

[62] Lena Reinfelder, Robert Landwirth, and Zinaida Benenson. 2019. Security Managers Are Not The Enemy Either. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19).* Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3290605.3300663

[63] Andrew Roberts, Olaf Maennel, and Nikita Snetkov. 2021. Cybersecurity Test Range for Autonomous Vehicle Shuttles. In *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW).* 239–248. https://doi.org/10.1109/EuroSPW54576.2021.00031

[64] Enrico Russo, Gabriele Costa, and Alessandro Armando. 2020. Building next generation Cyber Ranges with CRACK. *Computers & Security* 95 (2020), 101837. https://doi.org/10.1016/j.cose.2020.101837

[65] Martina Angela Sasse, Sacha Brostoff, and Dirk Weirich. 2001. Transforming the 'weakest link'—a human/computer interaction approach to usable and effective security. *BT technology journal* 19, 3 (2001), 122–131.

[66] Ahmad Zia Sharifi, Vajirasak Vanijja, Debajyoti Pal, and Watcharee Anantasabkit. 2021. CyberIoT: An Initial Conceptualization of a Web-based Cyber Range for IoT. In *2021 International Conference on Computational Performance Evaluation (ComPE).* 091–096. https://doi.org/10.1109/ComPE53109.2021.9752401

[67] Siddhant Shrivastava, Francisco Furtado, Mark Goh, and Aditya Mathur. 2022. The Design of Cyber-Physical Exercises (CPXS). In *2022 14th International Conference on Cyber Conflict: Keep Moving! (CyCon)*, Vol. 700. 348–366. https://doi.org/10.23919/CyCon55549.2022.9811000

[68] Michail Smyrlis, Konstantinos Fysarakis, George Spanoudakis, and George Hatzivasilis. 2020. Cyber Range Training Programme Specification Through Cyber Threat and Training Preparation Models. In *Model-driven Simulation and Training Environments for Cybersecurity*, George Hatzivasilis and Sotiris Ioannidis (Eds.). Springer International Publishing, Cham, 22–37.

[69] Iason Somarakis, Michail Smyrlis, Konstantinos Fysarakis, and George Spanoudakis. 2020. Model-Driven Cyber Range Training: A Cyber Security Assurance Perspective. In *Computer Security*, Apostolos P. Fournaris, Manos Athanatos, Konstantinos Lampropoulos, Sotiris Ioannidis, George Hatzivasilis, Ernesto Damiani, Habtamu Abie, Silvio Ranise, Luca Verderame, Alberto Siena, and Joaquin Garcia-Alfaro (Eds.). Springer International Publishing, Cham, 172–184.

[70] Dat Tang, Cuong Pham, Ken-ichi Chinen, and Razvan Beuran. 2017. Interactive cybersecurity defense training inspired by web-based learning theory. In *2017 IEEE 9th International Conference on Engineering Education (ICEED).* 90–95. https://doi.org/10.1109/ICEED.2017.8251171

[71] Zhihong Tian, Yu Cui, Lun An, Shen Su, Xiaoxia Yin, Lihua Yin, and Xiang Cui. 2018. A Real-Time Correlation of Host-Level Events in Cyber Range Service for Smart Campus. *IEEE Access* 6 (2018), 35355–35364. https://doi.org/10.1109/ACCESS.2018.2846590

[72] Elochukwu Ukwandu, Mohamed Amine Ben Farah, Hanan Hindy, David Brosset, Dimitris Kavallieros, Robert Atkinson, Christos Tachtatzis, Miroslav Bures, Ivan Andonovic, and Xavier Bellekens. 2020. A review of cyber-ranges and test-beds: Current and future trends. *Sensors* 20, 24 (2020), 7148.

[73] Stanislav Vakaruk, Alberto Mozo, Antonio Pastor, and Diego R. López. 2021. A Digital Twin Network for Security Training in 5G Industrial Environments. In *2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI).* 395–398. https://doi.org/10.1109/DTPI52967.2021.9540146

[74] Komal Bhupendra Vekaria, Prasad Calyam, Songjie Wang, Ramya Payyavula, Matthew Rockey, and Nafis Ahmed. 2021. Cyber Range for Research-Inspired Learning of "Attack Defense by Pretense" Principle and Practice. *IEEE Transactions on Learning Technologies* 14, 3 (2021), 322–337. https://doi.org/10.1109/TLT.2021.3091904

[75] Manfred Vielberth, Fabian Böhm, Ines Fichtinger, and Günther Pernul. 2020. Security Operations Center: A Systematic Study and Open Challenges. *IEEE Access* 8 (2020), 227756–227779. https://doi.org/10.1109/ACCESS.2020.3045514

[76] Manfred Vielberth, Magdalena Glas, Marietheres Dietz, Stylianos Karagiannis, Emmanouil Magkos, and Günther Pernul. 2021. A Digital Twin-Based Cyber Range for SOC Analysts. In *Data and Applications Security and Privacy XXXV*, Ken Barker and Kambiz Ghazinour (Eds.). Springer International Publishing, Cham, 293–311.

[77] Valdemar Švábenský, Jan Vykopal, Milan Cermak, and Martin Laštovička. 2018. Enhancing Cybersecurity Skills by Creating Serious Games. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (Larnaca, Cyprus) *(ITiCSE 2018).* Association for Computing Machinery, New York, NY, USA, 194–199. https://doi.org/10.1145/3197091.3197123

[78] Valdemar Švábenský, Richard Weiss, Jack Cook, Jan Vykopal, Pavel Čeleda, Jens Mache, Radoslav Chudovský, and Ankur Chattopadhyay. 2022. Evaluating Two Approaches to Assessing Student Progress in Cybersecurity Exercises. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1* (Providence, RI, USA) *(SIGCSE 2022).* Association for Computing Machinery, New York, NY, USA, 787–793. https://doi.org/10.1145/3478431.3499414

[79] Jan Vykopal, Martin Vizvary, Radek Oslejsek, Pavel Celeda, and Daniel Tovarnak. 2017. Lessons learned from complex hands-on defence exercises in a cyber range. In *2017 IEEE Frontiers in Education Conference (FIE).* 1–8. https://doi.org/10.1109/FIE.2017.8190713

[80] Jan Vykopal, Pavel Čeleda, Pavel Seda, Valdemar Švábenský, and Daniel Tovarňák. 2021. Scalable Learning Environments for Teaching Cybersecurity Hands-on. In *2021 IEEE Frontiers in Education Conference (FIE).* 1–9. https://doi.org/10.1109/FIE49875.2021.9637180

[81] Muhammad Mudassar Yamin and Basel Katt. 2022. Modeling and executing cyber security exercise scenarios in cyber ranges. *Computers & Security* 116 (2022), 102635. https://doi.org/10.1016/j.cose.2022.102635

[82] Muhammad Mudassar Yamin and Basel Katt. 2022. Use of cyber attack and defense agents in cyber ranges: A case study. *Computers & Security* 122 (2022), 102892. https://doi.org/10.1016/j.cose.2022.102892

[83] Muhammad Mudassar Yamin, Basel Katt, and Vasileios Gkioulos. 2020. Cyber ranges and security testbeds: Scenarios, functions, tools and architecture. *Computers & Security* 88 (2020), 101636.

[84] Muhammad Mudassar Yamin, Basel Katt, and Mariusz Nowostawski. 2021. Serious games as a tool to model attack and defense scenarios for cyber-security exercises. *Computers & Security* 110 (2021), 102450. https://doi.org/10.1016/j.cose.2021.102450

[85] Keiichi Yonemura, Hideyuki Kobayashi, Jun Sato, Hisashi Taketani, Shinya Oyama, Satoru Yamada, Satoru Izumi, Hiroyuki Okamoto, Youichi Fujimoto, Yoshinori Sakamoto, Kentaro Noguchi, and Seiichi Kishimoto. 2021. Cybersecurity Teaching Expert Development Project by KOSEN Security Educational Community. In *2021 IEEE Global Engineering Education Conference (EDUCON).* 468–477. https://doi.org/10.1109/EDUCON46332.2021.9453958

[86] Keiichi Yonemura, Shinya Oyama, Hideyuki Kobayashi, Satoru Yamada, Keiichi Shiraishi, Satoru Izumi, Tatsuki Fukuda, Hiroyuki Okamoto, Manabu Hirano, Youichi Fujimoto, Hideaki Moriyama, Yoshinori Sakamoto, Jun Sato, Kentaro Noguchi, Hisashi Taketani, and Seiichi Kishimoto. 2022. Teaching Expert Development Project by KOSEN Security Educational Community. In *2022 IEEE Global Engineering Education Conference (EDUCON).* 1643–1651. https://doi.org/10.1109/EDUCON52537.2022.9766560

[87] Alessandro Zanasi, Daniele Cristofori, and Graziano Giorgi. 2021. The European Commission contribution to cybersecurity through the ECHO project. In *2021 14th CMI International Conference - Critical ICT Infrastructures and Platforms (CMI).* 1–8. https://doi.org/10.1109/CMI53512.2021.9663786

[88] Verena Zimmermann and Karen Renaud. 2019. Moving from a 'human-as-problem" to a 'human-as-solution" cybersecurity mindset. *International Journal of Human-Computer Studies* 131 (2019), 169–187.

A APPENDIX: *TARGET* TAXONOMY

Table 5: *TARGET* Taxonomy: Learning Impact.

No.	Goal	References
LI.1	Willingness for continuous education and training	[25, 57]
LI.2	Projection of skills to real-world scenarios	[19, 21, 32]
LI.3	Long-term change in behavior	[32]
LI.4	Organizational impact	[67, 86]
LI.5	Societal impact	⊕

Table 6: *TARGET* Taxonomy: Learning Outcome.

No.	Goal	References
LO.1	Offensive security knowledge	[13, 29, 46, 67, 76, 77, 86]
LO.2	Offensive security skills	[1, 6–8, 25, 32, 33, 38, 41, 46, 55–57, 60, 61, 64, 67, 71, 77, 79, 81, 82, 84, 85]
LO.3	Defensive security knowledge	[29, 67, 76, 77, 86]
LO.4	Defensive security skills	[1, 3, 7, 8, 19, 26, 29, 32, 42, 51, 56, 57, 60, 61, 67–69, 74, 76, 79, 81, 82, 84–86]
LO.5	Forensic knowledge	[28]
LO.6	Forensic skills	[7, 8, 28, 42, 60, 71, 82, 84]
LO.7	Security awareness	[33, 38, 42, 57, 70]
LO.8	Non-security related technical knowledge	[29, 57, 77]
LO.9	Non-security related technical skills	[38, 57, 67, 77]
LO.10	Skills for coping with stressful situations	[21, 28, 51]
LO.11	Teaming/collaboration skills	[8, 21, 22, 38, 51, 77]
LO.12	Teaming/collaboration knowledge	[77]
LO.13	Decision-making skills	[38, 68, 77, 84]

Table 7: *TARGET* Taxonomy: Learning Experience.

No.	Goal	References
LE.1	Suitability for target group	[28, 60, 68, 79]
LE.2	Engagement of trainees	[25, 29, 32, 45, 51, 76, 79]
LE.3	Adequacy of difficulty	[29, 32, 51, 77, 81]
LE.4	Adequacy of duration	[66, 70]
LE.5	Clarity of tasks	[70]
LE.6	Quality of learning material	[86]

Table 8: *TARGET* Taxonomy: Teaming.

No.	Goal	References
T.1	Diversity of different teams	[1]
T.2	Dynamic teaming roles configuration	[50]
T.3	Balance of skills among teams	[30, 40, 51, 77, 79]
T.4	Team effectiveness	[11, 30, 40, 77]
T.5	Leadership effectiveness	[11]

Table 9: *TARGET* Taxonomy: Scenario.

No.	Goal	References
S.1	Relevance	[12, 46, 67, 70, 77, 81, 82]
S.2	Currency	[8, 67]
S.3	Fidelity of attacks	[3, 69, 74]
S.4	Fidelity of simulation/emulation	[21, 22, 26, 33, 34, 50, 51, 56, 57, 60, 63, 64, 69, 71, 73, 74]
S.5	Fidelity of tools	[19, 28, 56, 76]
S.6	Customizability	[9, 26, 34, 42, 46, 66, 68, 69, 81, 81]
S.7	Adaptability of difficulty	[4, 46, 81]
S.8	Immediate feedback for trainees	[1, 21, 34, 55, 55, 79, 79]
S.9	Tool-supported collaboration	[46]
S.10	Alignment with certification requirements	[32]
S.11	Alignment with cybersecurity curricula	[77]

Table 10: *TARGET* Taxonomy: Management.

No.	Goal	References
MG.1	Cost-efficiency	[8, 9, 12, 25, 26, 29, 41, 49, 63, 66, 79]
MG.2	Structured deployment	[6, 9, 12, 13, 42, 50, 60, 74, 87]
MG.3	Structured scenario creation	[4, 6, 8, 9, 12, 21, 46, 50, 61, 64, 68, 80–82]
MG.4	Variety of training scenarios	[13, 26, 47, 53, 63, 66, 73, 74]
MG.5	Technical complexity of orchestration	[7, 66]
MG.6	Automation of attacks	[9, 19, 54, 61, 70, 73, 76, 79, 82]
MG.7	Automation of defensive security operations	[81, 82]
MG.8	Automation of event generation	[19, 31, 81]
MG.9	Automation of orchestration	[8, 54]
MG.10	Automation of training content generation	[7, 8, 54]
MG.11	Automation of deployment	[6, 31, 41, 45, 54, 64, 68, 70]
MG.12	Automation of testing	[64]
MG.13	Collaboration between different cyber ranges	[6, 34, 46, 49, 87]
MG.14	Collaboration with the industry	[85, 86]

Table 11: *TARGET* Taxonomy: Monitoring.

No.	Goal	References
MO.1	Automation of infrastructure monitoring	[6, 53, 61, 82]
MO.2	Automation of trainee monitoring	[10, 31, 50, 55, 61]
MO.3	Sophistication of trainee scoring	[10, 46, 61, 67, 81]
MO.4	Visibility	[45, 80]

Table 12: *TARGET* Taxonomy: Environment.

No.	Goal	References
E.1	Accessibility	[38, 66, 67, 70, 74, 76, 79, 80]
E.2	Accuracy	[81]
E.3	Resource-efficiency	[12]
E.4	Reliability	[9, 51, 53, 67, 77, 81]
E.5	Compatibility with standard technologies	[21, 64]
E.6	Extensibility	[64, 74]
E.7	Independence from deployment infrastructure	[9, 80]
E.8	Isolation	[12, 34, 79]
E.9	Modularity	[13, 42, 47]
E.10	Open-source availability	[42, 60, 66, 79]
E.11	Performance	[7–9, 12, 13, 47, 51, 53, 54, 71, 81]
E.12	Portability	[13, 46, 47, 73]
E.13	Scalability	[9, 21, 26, 45, 49, 50, 53, 56, 60, 64, 66, 69, 73, 80, 81]
E.14	Reproducibility of scenario	[7, 12, 13, 47, 63, 73, 77, 82]
E.15	Reusability of scenario (components)	[6, 8, 28, 46, 60, 64, 80, 87]
E.16	Security	[1, 6, 34, 47, 53, 60, 66]
E.17	Privacy	⊕