Learning Analytics Perspective: Evidencing Learning from Digital Datasets in Cybersecurity Exercises

Kaie Maennel
School of Information Technologies: Department of Software Sciences
Tallinn University of Technology
Tallinn, Estonia
kaie.maennel@taltech.ee

***Abstract*—Cybersecurity exercises are gaining in popularity in university curricula and professional training paths and are seen as an effective teaching method. Such exercises provide digital datasets to facilitate a learning analytics approach such as by using the traces that learners leave behind to improve the learning process and environment. While there are various learning measurement efforts from digital datasets in the existing literature, a holistic learning analytics approach incorporated into cybersecurity exercises is still lacking. We propose a practical reference model for incorporating a learning analytics approach into the cybersecurity exercise life-cycle. To facilitate this application, we have performed an extensive review of existing academic research on applying learning analytics in the context of cybersecurity exercises. We specifically focus on the learning indicators used to measure empirical impact and training effectiveness that could indicate achievement of defined learning outcomes. This reference model and overview of existing learning analytics use cases and learning metrics in various types of exercises can help educators, organisers and cyber range developers. This results in more adaptive exercise design and measurement using evidence-based data and connects digital learning traces to skills and competencies.**

***Index Terms*—cybersecurity, learning analytics, learning metrics, training, exercises**

1. Introduction

Effective learning, teaching, and skills improvement of cybersecurity students and professionals is a critical research area. As there is a high demand for skilled professionals and a shortage of such individuals [1], the development of scalable and effective teaching methods is critical. We focus on the application of learning analytics in the cybersecurity training, specifically in cybersecurity exercises, as a way to provide a more evidence-based and systematic approach for the evaluation of learning impact and to enable the design of more effective learning. This is a critical aspect to consider for educators, organisers and cyber range developers.

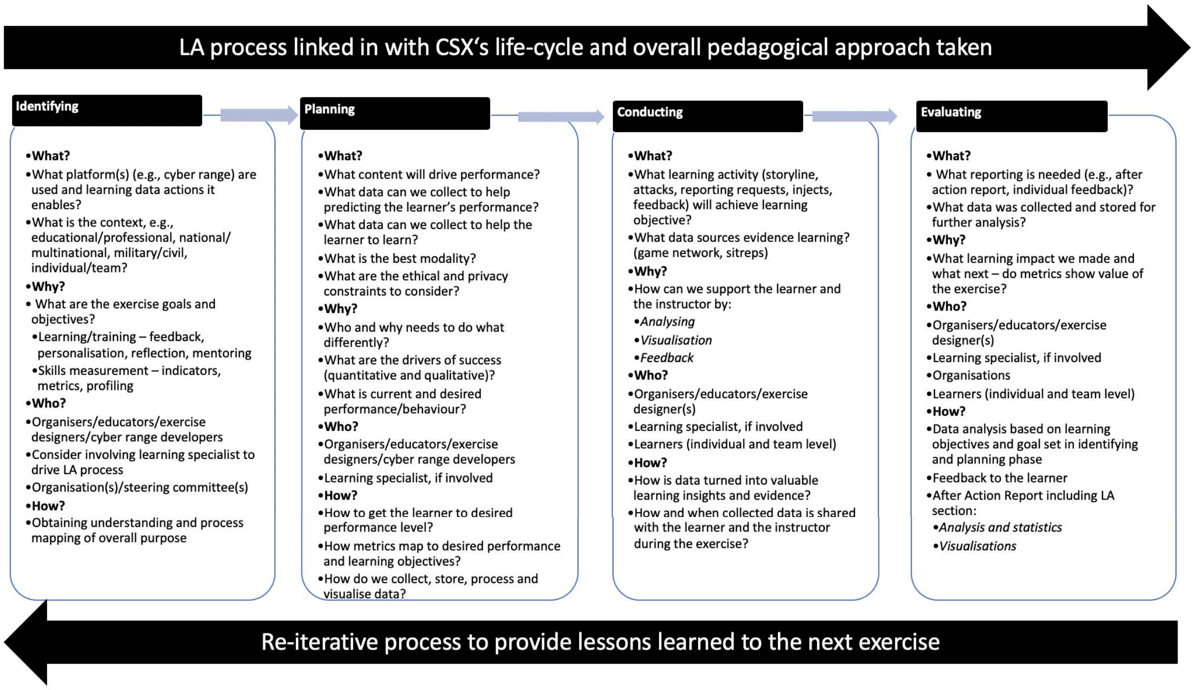
Learning analytics (LA) is defined as “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” [2]. As a field of research, LA aims to predict and advise on identifying students’ learning needs and improve pedagogical strategies based on analytical approaches [2]. However, establishing plausible relationships between models derived from quantifiable digital data, and the complex socio-cognitive world of “learning” is challenging [3]. LA is closely intertwined with educational data mining (EDM) that develops, researches, and applies computerized methods to detect patterns in large educational data sets [4].

Cybersecurity training teaches both technical and soft skills, as the field involves technology, people, information, and processes. A wide range of training methods have been developed by universities [5] and organisations to provide cybersecurity education. As part of such cybersecurity trainings, hands-on exercises (both online and classroom) are gaining in popularity in university curricula and professional training paths. Cybersecurity exercises (CSXs) are viewed as an effective and engaging way of teaching both technical and soft skills in addition to CSXs for learning purposes (e.g., as part of university courses, competitions across universities, etc.). Most national and international CSXs (47%) also focus on training and providing participants an opportunity to gain knowledge, understanding and skills [6]. The CSXs can vary significantly in scale and content, ranging from short online or classroom exercises, Capture the Flags (CTFs) to large-scale/multi-stakeholder exercises, etc. However, most share common aims and challenges with respect to learning. We view CSX as a learning or training event in which individuals or teams implement, manage and defend/attack a network of computers at a tactical or strategic level.

As those exercises always leave an extensive digital footprint of learning processes, it makes them an ideal base to develop the methods within the learning analytics field itself. As a result, using these evidence-based learning traces in learning design can improve the experience for both students and specialists. It also helps to investigate the validity of common, yet unsubstantiated claims, such as “everyone feels they had learned important lessons [7]” or “exercises are a very effective way of learning the practical aspects of information security” [8].

We propose a reference model for LA in CSXs in Section 2 offering a practical guide to the exercise organisers to enhance the conceptualisation and integration of learning analytics into the exercise life-cycle. We support the proposals presented in the model with an extensive overview of existing uses of learning analytics by pro-

DOI 10.1109/EuroS&PW51379.2020.00012


LA process linked in with CSX‘s life-cycle and overall pedagogical approach taken
Identifying

- **What?**
- What platform(s) (e.g., cyber range) are used and learning data actions it enables?
- What is the context, e.g., educational/professional, national/multinational, military/civil, individual/team?
- **Why?**
- What are the exercise goals and objectives?
  - Learning/training – feedback, personalisation, reflection, mentoring
  - Skills measurement – indicators, metrics, profiling
- **Who?**
- Organisers/educators/exercise designers/cyber range developers
- Consider involving learning specialist to drive LA process
- Organisation(s)/steering committee(s)
- **How?**
- Obtaining understanding and process mapping of overall purpose

- **What?**
- What content will drive performance?
- What data can we collect to help predicting the learner’s performance?
- What data can we collect to help the learner to learn?
- What is the best modality?
- What are the ethical and privacy constraints to consider?
- **Why?**
- Who and why needs to do what differently?
- What are the drivers of success (quantitative and qualitative)?
- What is current and desired performance/behaviour?
- **Who?**
- Organisers/educators/exercise designers/cyber range developers
- Learning specialist, if involved
- **How?**
- How to get the learner to desired performance level?
- How metrics map to desired performance and learning objectives?
- How do we collect, store, process and visualise data?

- **What?**
- What learning activity (storyline, attacks, reporting requests, injects, feedback) will achieve learning objective?
- What data sources evidence learning? (game network, sitreps)
- **Why?**
- How can we support the learner and the instructor by:
  - *Analysing*
  - *Visualisation*
  - *Feedback*
- **Who?**
- Organisers/educators/exercise designer(s)
- Learning specialist, if involved
- Learners (individual and team level)
- **How?**
- How is data turned into valuable learning insights and evidence?
- How and when collected data is shared with the learner and the instructor during the exercise?

- **What?**
- What reporting is needed (e.g., after action report, individual feedback)?
- What data was collected and stored for further analysis?
- **Why?**
- What learning impact we made and what next – do metrics show value of the exercise?
- **Who?**
- Organisers/educators/exercise designer(s)
- Learning specialist, if involved
- Organisations
- Learners (individual and team level)
- **How?**
- Data analysis based on learning objectives and goal set in identifying and planning phase
- Feedback to the learner
- After Action Report including LA section:
  - *Analysis and statistics*
  - *Visualisations*

Re-iterative process to provide lessons learned to the next exercise

Figure 1. CSX Learning Analytics Reference Model

viding empirical evidence from digital datasets (log files, pcaps) and metrics used in CSXs.

When implementing LA measurements into exercises we need to understand (1) what metrics evidence learning and (2) are they helping the learners to learn or teachers to teach? The metrics (i.e., indicators of learning success) collected and analysed provide technical data (e.g., time, command-line, tools used, etc.), and there appears to be an overemphasis upon what we can measure, instead of measuring what we value—a longstanding concern in educational assessment [9]. However, applying a learning analytics approach and analysing metrics from digital datasets, can provide a more detailed and evidence-based input to more comprehensive learning evaluations, such Kickpatrick or other chosen evaluation models [10].

2. Reference Model for LA in CSXs

Learning analytics should be incorporated to the CSXs' identifying, planning, conducting and evaluating phases (as described by [11]) and be seen as an integral part of the exercise design in line with the overall pedagogical approach selected [12]. When starting to implement a LA approach into an exercise it is useful to think about LA process from aspects of What (Data, Environments, Context), Why (Objectives), Who (Stakeholders) and How (Methods) [13].

We propose a practical tool: the CSX LA reference model, Fig. 1, that builds upon [13] and [11]. Our contribution is to combine and outline the key learning analytics considerations to incorporate into the CSXs life-cycle and support the model with an extensive overview of existing use cases for a practical implementation. Developers would need to consider LA aspects in their initial design of the cyber ranges when they incorporate the technological foundation of instrumenting the exercises.

Asking these learning analytics related questions and finding the answers during an exercise life-cycle, will ensure that learning measurements are not simply an afterthough but rather are incorporated in the "identifying phase" (Fig. 1). Considering questions, such as "What data can we collect that will help learners to learn?" and collecting only that relevant dataset, would help with the challenges of storing huge datasets from an exercise and later trying to determine what data could be used to provide feedback. For example, if the learning objective of an exercise is to improve the incident response process, then timestamps that would indicate team communication would be critical data to collect [14]. However, when the proficiency of using various forensics tools and commandlines is exercised, then capturing bash history or keystrokes is relevant (e.g., [15], [16]). Also consideration should be given how to support instructors in giving feedback. Designing automated feedback that takes into account the users behaviour and predicts their actions and questions becomes available and can make the learning experience more individualised and effective (e.g., [17], [18]). Depending on the purpose, scale and type of a CSX, it may be recommended to include a learning specialist in the organising team to coordinate the LA implementation throughout the CSX life-cycle.

2.1. Comparison to Other Frameworks in CSXs

There are various frameworks that have been developed for CSXs and cybersecurity education. Several academic papers (e.g., [8], [56]) and non-academic guides (e.g., [11], [57], [58]) describe the overall CSXs de-

TABLE 1. Mapping LA processes by types of CSXs. Note: One paper may cover various topics, we map here main LA theme.

Exercise Type/LA Process	Collection	Storage	Cleaning	Integration	Analysis	Visualisation	Action
Capture the flag	[19], [20], [21], [22], [23], [24], [25]	[23]		[22]	[19], [20], [21], [22], [26]	[27], [28]	
Discussion based game							
Drill							
Red team / blue team	[29], [14], [30], [31]	[32]			[14]	[33], [33], [34]	
Seminar							
Simulation	[35], [36], [37], [38], [39], [40], [41], [42]			[35]	[35], [36], [43], [44]	[38], [39]	
Table-top	[45], [46]					[45]	
Workshop	[47], [48]						
Exercise/lab	[15], [49] [50], [16], [51], [52], [53], [8], [54]		[50]	[52], [53]	[50], [55]	[15], [49], [32]	[17]

sign and evaluation process. Also, more general approaches to exercises are proposed: [59] that describe an extended competence development and assessment framework and [56] suggests specific metrics in complex simulated CSXs. For designing competition based exercises [60] describes a mindmap, while for CTFs [61], [62] describe 5 steps when designing an evaluation (purpose, frame, questions, information needed and systematic collection method). At a higher level conceptual level, models for a multidisciplinary cybersecurity training methodology [63], a pedagogical framework [12], a framework incorporating cognitive aspects [64], and a holistic model of professional competence in the cyber domain [65] have also been developed.

Despite some of these models providing elements of analytical approaches, no framework has been developed that would explicitly include the use of LA methods in the CSXs when looking at the LA in each phase of exercise life-cycle or what would be the appropriate metrics or learning indicators to measure when considering the pedagogical approach taken.

3. Supporting Reference Model with Existing Research Results and Practical Considerations in Implementing LA Approach in CSXs

To gain an insight about what empirical evidence and metrics have been collected and analysed in the scientific literature, we conducted extensive related work review. We searched Google Scholar, a widely used and available scientific database and limited the search to empirical studies published in a peer-reviewed journals and conferences in English. We used the keyword "learning" in combination with "cyber(-)security" for different exercise types. For the exercise type classification we followed the European Union Agency for Network and Information Security (ENISA) taxonomy: Capture the flag, Discussion based game, Drill, Red team / Blue team, Seminar, Simulation, Table-top and Workshop [6]. In some cases, exercises with gamified elements are referred as "serious games", or with competitive elements as "competition", and thus included these in our search strings for completeness.

We reviewed the abstracts of 200 articles for each exercise type, as this was considered sufficient to encompass all relevant material. The academic papers identified as covering CSXs for learning purposes were manually reviewed to identify LA topics and all empirical/analytical learning data collected or analysed. It should be noted that even though there is a large number of articles describing the CSXs, these do not necessarily include empirical data from the digital datasets to evidence learning. We used feature mapping [66] in which the content is analysed and recorded in a standardized format documenting the key features of a predetermined aspects (i.e., LA model [2]) to produce a summary of the topic. Related work was mapped to an overview matrix in Table 1 by the ENISA exercise types [6] and LA model [2] consisting of collection, storage, data cleaning, integration, analysis, representation and visualization and action.

3.1. Uses of Digital Datasets to Evidence of Learning Effectiveness in CSXs

An overview matrix that shows the application of learning analytics by exercise types and LA process steps is presented in Table 1.

3.1.1. Capture the Flag. A CTF is typically a challenge designed to help sharpen cybersecurity skills and provide hands-on learning taking various styles, such as jeopardy, attack-defence and a mix of the two. However, participating in CTFs does not necessarily ensure future success, and participants rarely receive a detailed critique of their performance, which is essential in learning [67].

There are few studies that provide empirical evidence from digital game-play traces for learning and skill acquisition in CTFs. Clothia and Novakovic [19] show that Jeopardy-style challenges with automatic marking of flag submissions complemented by manual marking of detailed written answers provided students with instant feedback during an exercise with an improved student satisfaction with the academic course. The ability of students to acquire the flag is highly correlated with their overall marks (written assignment), and flag-based marking effectively assesses a student's basic skills and understanding of cybersecurity topics [19]. However, acquiring flags and a student's deeper understanding of the underlying issues is much less-correlated [19]. Cheung et al. [20] describe combining logging results from log servers along with a key-logger to track participant sessions. The authors gathered statistics about login times (e.g., the environment was most often used on Saturdays between 3:30pm and 8pm) and command usage to see what commands students were still having trouble with after the lectures, which helped determining what to spend more time on [20]. Chapman et al. [21] evaluate PicoCTF based on survey responses and user interaction logs to explore the effectiveness of design choices (e.g., younger students prefer a game interface compared to older students, and a general dislike for challenges requiring learning new tools). The system kept track of answers submitted by every team, both correct and incorrect—recording the time, content and relevant problem identifier, as well as the IP address of the submission [21]. In order to evaluate student engagement, the authors determined periods of time during which teams were most active (i.e., time interval at submissions) [21]. Tseng et al. [22] also focused on data collection from heterogeneous environments and proposed an ontology to represent concepts (consisting of teams profile, skills (tools), test items and environment) within exercises and their relationships (including linking heterogeneous logs to participants' intentions). The authors focused on problem-solving behaviours and applied a modified a priori algorithm, analysed frequent item-sets and identified learning behaviours such as that novices keep guessing keys and better performing students focused on particular items [22]. The above papers describe metrics applied and the authors research selection of patterns and relations in the digital learning dataset. [23] describe possible metrics, such as teamwork, challenge difficulty, challenge strategies, tools, and general problem solving techniques. Whereas [27] and [26] propose, apply, and experimentally evaluate data analysis and machine learning techniques to obtain interactions from the in-game data and provide learners who progress differently with individualized feedback. Vykopal et al. [25] suggest decomposing the exercise training activity into individual levels to achieve specific learning objectives, and collecting timestamps (or events) such as start and end of the game, start and end of each level, submission of incorrect flags and their content, hints used, skipping a level, displaying a level's solution, game ID [24].

Dark and Mirkovic [61] bring out an aspect that to measure learning we may need to rely on proxy indicators (e.g., identifying reasonable and observable indicators of adversarial thinking). Overall, we can see that analytical approaches are emerging to evaluate learning from digital dataset(s) that are good starting points when incorporating the learning analytics process into the exercises.

3.1.2. Red Team and Blue Team. Red-Blue exercises are often team-based and therefore add another complexity level in LA application—measuring team learning vs. individual learning. Several authors discuss learning impact; however, not evidence-based analysis from digital footprints (e.g., [68] describes organizing team-based exercise, where teams were directly monitored and evaluation of skills improvement is observational). Typical evaluation methods are score-boards, verbal feedback and after-action reports highlighting conclusions from manual analysis of exercise data [34] that do not apply an analytical approach using digital datasets. Some papers do describe data collection and perform initial LA on digital dataset, e.g., [14] breaking time into intervals that can be meaningful for different learning objectives (e.g., incident responding, team communication) with walk-through. Ošlejšek et al. [33] show that visual analytics tools could provide automated statistical analysis and an in-depth insight into the learner's behaviour using observation software (Fowler's analysis) [32]. Vykopal et al. [34] describe an interactive timeline visualisation allowing learners to explore a scoring timeline and details about individual events.

3.1.3. Simulation. Simulations are a very common type of exercise and take various forms, e.g., online vs. live, gamified, etc. There are some emerging examples of LA performed already. Thompson and Irvine [35] present basic LA data, such as time played and discuss abstraction layers for data analysis. No formal effectiveness assessment was performed and conclusions are based on ad-hoc student interaction and logs review [35]. Legato and Mazza [36] assume a set of regeneration points that correspond to skill achievement through learning. However, the model was not validated and numerical results are reported for illustrative purposes only.

Nicholson et al. [37] system uses a dynamic tailoring system, which maintains a model of student proficiency and adapts training difficulty, while providing detailed feedback. Santos et al. [38] use a post-simulation analysis using a variety of graphics and reports to verify the network traffic, which teams were attacked, which services are still vulnerable, teams activity rate and strategy used, etc. This data enables statistical evaluation of what happened during the simulation, including how teams perform compared to previous exercises [38]. Furfaro et al. [39] used cloud based learning system including a dashboard (for managing scenarios, agents and VMs, displaying system usage and statistics, etc.); a report tool (provides statistical data from logging engine, and queries for business intelligence analysis displayed on charts), and a set of development tools.

Many simulations have a gamification element. Tioh et al. [69] performed a literature review (18 papers including some kind of empirical effect measurement) and concluded "the question as to the effectiveness of serious games dealing with the training of cyber security is a difficult one to answer conclusively at this point". We identified additional papers with games for security specialists or students (e.g., [69], [40], [41]). However, none used empirical learning analysis from digital datasets.

Several authors focus on cognitive levels of learning process in articles covering simulation-type exercises, as simulations allow experimentation. Some examples include a simulation-based approach to understanding cybersecurity threats when attempting multiple actions, the user is provided with an "awareness" measure [70]; a socio-technical systems approach to support the emerging role of systems thinking and using an agent-based simulation tool to change the students' thinking [71] [72]; a computational model based on Instance-Based Learning Theory that proposes a way to analyse the cyber analyst's awareness at both threat level and attack scenario level [43]. Such human dynamic decision making analysis can help to determine various player models at individual and aggregated levels [73].

3.1.4. Table-top. A table-top exercise is typically a meeting to discuss a fictional cyber emergency situation increasing participants' engagement and strengthening their awareness and competences in strategic decision-making [74]. Although typically digital traces are typically limited in table-tops, some research on the system architecture for tracking learning process is emerging. Brilingaite et al. [45] presents a model of a web-based environment that enables playing table-top exercises in person and remotely. The environment includes the visual representation of decision-making during the game and provides the comparison to the correct solution.

However, in many cases the data analysis is not presented, offering an experience without learning process or empirical impact data. Cozine et al. [75] examine the pedagogical approach to incorporate game play, specifically probability-based tabletop exercises, into course curricula and collected survey data from students enrolled in the courses. Ottis [46] presents a light weight tabletop exercise format that has been successfully used in cybersecurity education to demonstrate these and many other concepts to master level students.

3.1.5. Drill, Seminar, Discussion Based Game and Workshop. The research builds upon experience but lacks evidence-based measures or uses mainly surveys/self-assessments as a tool to analyse learning. Some research results start to emerge, such as [47] describing a workshop using clickers for running a series of questions that allow easy data collection and analysis.

3.1.6. Exercises with no clear classification. In several papers CSXs are described in generic manner as "exercise" (often in classroom environment and part of a course) or include multiple exercise types (e.g., platform allowing both Red-Blue team and individual simulation game). However, several empirical learning data analyses have been completed. Weiss et al. [15] express that simply recording the number of correct answers is inferior to in-depth assessments and explores the use of command line history and visualization. Authors follow the "path" taken by a student in command-line when completing different tasks and levels (for skills level measurement some commands were identified as significant) [15]. The "path" is visualized in graph that can be decomposed into chains and cycles [15]. Similarly, Labushange et al. [49] assesses technical skill level based on indexed similarity (i.e., participants were ranked based on commands usage to achieve objectives) and classifies actions that can be automatically deducted using the clustering of commands (e.g., combination of "ifconfig", "sudo apt-get install nmap" and "sudo nmap -sT targetIP" together was classified as reconnaissance). However, the paper does not go into details of how such clustering can be achieved [49]. Caliskan et al. [50] use educational data mining, machine learning and identifies metrics for learning effectiveness (predicting final grade) in the university classroom course with efforts to validate their predictive model. Caliskan et al. also compare participant evaluation metrics and scoring systems in [76]. Moore et al. [55] focuses on the development of specific individual skill levels and state that competence in progressively harder levels of capabilities was observed over time in relation to the training components. [52] apply automated mechanism for parsing log entries into blocks of time during which participants are focused on specific high-level objectives, with instrumentation capturing students' computer-based transactions [53]. Several authors propose evaluation metrics or data that should be collected. However, the actual learning data analysis is lacking so far. For example, [54] suggest metrics such as time, participant numbers who succeed and feedback, and [8] suggests number of detected attacks from total attacks for learning task of monitoring systems' security, etc.

3.2. Analysis of LA Process Described in CSXs

The explicit use of learning analytics and relevant vocabulary in the cybersecurity education (incl. CSXs) is in its early stages. However, recently there is more focus in the academic community on applying LA methods to improve the cybersecurity education [77]. The discussion below is organized by LA process to analyse main steps in a CSX's life-cycle. The process steps span across the life-cycle and are re-iterative, however the whole process needs to be designed in planning phase and instrumented to the cyber ranges.

3.2.1. Collection and Acquisition. Existing research focuses on data collection—i.e., how to build an exercise platform, cyber range, etc. However, the literature lacks considerations for what purpose and what data is actually collected, and also contains discussion on learner consent and ethical aspects. As the tendency is for collecting simple technical measures, rather than more complex cognitive learning measures (Section 4), often there is no clear connection whether it was collected for evidencing learning, and what data is relevant for evidencing learning.

3.2.2. Storage. How and what data is stored (and how long period) is mostly not covered (only few examples such as [32]). The security of information stored and privacy concerns (anonymization processes) appear not to be a high priority. Only one paper [23] was identified that describes security measures for captured data.

3.2.3. Cleaning. The data cleaning process is not typically described, however network traffic (pcap) and other datasets are expected to include non-relevant data. Few papers start discussing what processes were used to clean the data, e.g., [50]. Also the principles of privacy and anonymization needs to considered here.

3.2.4. Integration. As the exercises generate multiple datasets, combining multiple datasets and different formats, e.g., technical and timing data with self-reported learner data, is an important process step. Some examples were found about aligning timestamps of datasets, such as [52], [53] [14], [22]—however, no detailed methods are provided on how data is processed and time-synced.

3.2.5. Analysis. Due to LA being a novel research area, the data analysis performed has been limited, with a predominance of studies undertaking low-level analysis relating to the readily accessible data such as the reporting of number/frequency login times, number of messages posted, time online, etc to academic performance as measured by grades [78]. Similarly in CSXs, we did not identify commonly used tools or methods for data analysis (statistical, machine learning model, etc.). Similarly to overall LA field, the analysis has genreally been conducted using easily obtainable metrics (such as time), see Table 2 and in typically linked to high-level learning objective(s).

3.2.6. Representation and Visualization. The challenge is determining the relationship between visualizations and learning. The feedback about low level user actions—such as number of log ins, videos watched, or documents submitted—does not illustrate progress in learning for students or educators [79]. Visualizations and dashboards usefulness and effectiveness is not widely covered in the exercises. Several papers, such as [32] analyse the use of visualizations in CSXs in addition to describing the system architecture.

3.2.7. Action. Actions, such as intervention, optimization, systematic improvements (including design) are not necessarily evidence-based. Rather learning design choices are based on the authors' experience or the learners' self-reporting/survey evaluations. Some relevant research is emerging how to improve feedback loop from using digital traces, e.g. [17].

4. Inferring Learning from Digital Datasets—What to Measure?

The various frameworks in Section 2.1 have been developed for learning in CSXs and cover different aspects but none directly incorporate or utilize learning analytics processes. When inferring learning from the granular digital dataset, the challenge is linking learning objectives and competencies to the granular raw data to, as the design of a CSX should follow a top-down pattern [18]. The cyber range should be designed to allow such learning design and measurement process.

4.1. What Metrics are Collected and Analysed to Evidence Learning?

Table 2 summarises the learning indicators from digital datasets that have been used in academic research, which could be used as a starting point to brainstorm when selecting the metrics to measure that the training objectives have been achieved. It should be noted, that any papers on learning in the CSXs are based on the experience and interpretation of the authors or based on the traditional learner evaluation (e.g., feedback surveys, evaluation forms).

4.2. What to Consider in Choosing Metrics?

We should focus on measuring what we value. The metrics used in CSXs often focus on easily measurable data (e.g., time spent, number of attacks mitigated, etc.) and individual actions. However, the students are "too easily satisfied that a system is secure after identifying only one possible source of security for a system rather than seeking to explore the adversarial space more thoroughly" [83]. Thus it is important to understand not only whether the students found the correct answer but how they found it [15]. There is some research that starts to look into "how" the learner completes tasks (i.e., use of tools, attempts, submission of wrong answers), such as [52], [53], [49]. However, validation is limited (e.g., 4 participants [49]). In regards to teamwork and communication, there is some research, such as [81], [36] that have started to explore the use of analytics as evidence for achieving learning in teams.

Also as learning is complex cognitive process, the further research should focus on cognitive metrics, such as Knox et al. [82]. From the LA research, a similar measure to "cognitive presence" can be applied in cybersecurity training (e.g., "Active Learning Squared (AL2)" paradigm, which emphasises metacognition and uses both active student learning and machine learning [84], [85]).

Metrics are valuable, however, "being able to report upon a metric does not mean that you should use it, either in the tool, or in reporting its worth [3]". The metrics will depend on the exercise goals that in turn are guided by different pedagogical principles (e.g., behaviorist, cognitivist or constructivist) [12] and the wider evaluation model chosen [10]. Therefore, we need to be mindful of learner and learning process, and measurement should move towards mapping of digital traces describing student activity onto interpretable constructs of interest (e.g., Knowledge Components, Q-matrix), which facilitate actionable analytics [86].

5. Challenges in Implementing LA approaches in CSXs

Scientifically-valid evidence that learning outcomes were achieved in CSXs is difficult to obtain, especially as the exercise design, objectives, technology and learner characteristics vary. These factors make inter-institutional and between exercises comparisons difficult. However, sharing the measurement results would enhance measuring that the learning was achieved and new skills obtained.

The data analysis until now has been limited, with a predominance of studies undertaking low-level analysis using easily obtainable metrics, such as login times, time to complete tasks, number of attack mitigated, see Table 2. The related work did not reveal commonly used data analysis tools or methods (statistical, machine learning, etc.) in CSXs, but developing and sharing methods used would enhance validity of the results.

Security challenges, such as intrusion detection, insider threats, malware detection, and phishing detection lack exact algorithmic solutions and the boundary between normal and anomalous behaviour isn't clear-cut as attackers are continuously improving their techniques and strategies [87]. This also impacts LA, as it needs to keep up with moving algorithms and learning patterns. In addition, the challenge relates to data volume—one large exercise can create terabytes of data including multiple

TABLE 2. METRICS FROM DIGITAL DATASETS TO CONSIDER WHEN MEASURING LEARNING IN CSXS

Metrics	Reference	Learning Objective/Competency	Validated	Validation Method/Results
Technical Metrics				
Time and Time Periods				
Total completion time	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
Time taken to win the exercise	[54]	Effectiveness of the overall exercise	No	Recommendation
Time before nmap	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
Time spent on scenario (incl replays)	[35]	Network filters	No	Interpretation, CyberSiege, 1 lab, 149 students
Time taken to recover from a successful attack	[8]	Incident handling / response	No	Recommendation
Downtime of attacked service compared to attack duration	[8]	Perform DDoS	No	Recommendation
Time period during the attack response (5 timestamps)	[14]	Incident response / handling	No	Log analysis vs. self-reporting, Locked Shields, 19 teams
Time played	[35]	Various cybersecurity skills	No	CyberCiege, online platform
Mean time per action	[52], [53]	Forensics	No	Interpretation, TracerFire, 26 participants, 2 CSXs
Commands, including count of commands				
Time-to-Detect	[80]	Defending network against attack	No	Bivariate regression analysis, multivariate regression analysis, and principal component analysis
Time-to-apProval (by team controller)	[80]	Defending network against attack	No	Bivariate regression analysis, multivariate regression analysis, and principal component analysis
Time-to-End	[80]	Defending network against attack	No	Bivariate regression analysis, multivariate regression analysis, and principal component analysis
Category Correct (NIST category of inject correctly identified)	[80]	Defending network against attack	No	Bivariate regression analysis, multivariate regression analysis, and principal component analysis
Total Commands Entered	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
Reconnaissance Similarity (index of most accurate command)	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
Number of File Commands	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
File Server Identification Similarity (index of most accurate command)	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
Number of Incident Commands	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
Incident Commands Similarity (index of most accurate command)	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
Number of Threat Commands	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
Threat Commands Similarity (index of most accurate command)	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
Number of System Administration Commands	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
System Administration Similarity (index of most accurate command)	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
Command usage	[20]	Various cybersecurity knowledge	Partial	Interpretation, feedback survey, CTF with lectures, no details of flags or commands
Count of Events, Objects or Individuals				
Number of scenario replays	[35]	Network filters	No	Interpretation, CyberSiege, 1 lab, 149 students
Number of successful attacks	[8]	Implement security configurations	No	Recommendation
Number of detected attacks from total number of attacks	[8]	Monitor systems' security	No	Recommendation
Number of attacks correctly identified	[8]	Analyse logs and do forensics	No	Recommendation
Number of open ports/services detected compared to total number of open ports	[8]	Perform scanning and enumeration	No	Recommendations
Number of successful backdoors accesses to target systems kept until the exercise end	[8]	Cover tracks and place backdoors	No	Recommendations
Number Actions Per Block	[52], [53]	Forensics	No	Interpretation, TracerFire, 26 participants, 2 CSXs
Number Actions Per Block	[52], [53]	Forensics	No	Interpretation, TracerFire, 26 participants, 2 CSXs
Number of finalist participants	[54]	Effectiveness of exercise	No	Recommendation
Number of participants succeeding brute-force attack	[54]	Effectiveness of exercise	No	Recommendation
Number of participants successfully exploited Windows vulnerability	[54]	Effectiveness of exercise	No	Recommendations
Compromised services as reported by attacking and defending teams	[81]	Various cybersecurity skills	No	Statistical analysis (team performance)
Number of attack and vulnerability reports per defending team	[81]	Various cybersecurity skills	No	Statistical analysis (team performance)
Attempted and successful attacks on teams DMZ, calculated by NIDS analysis	[81]	Various cybersecurity skills	No	Statistical analysis (team performance)
Number Different Software Tools	[52], [53]	Forensics	No	Interpretation, TracerFire, 26 participants, 2 CSXs
Number Transitions Between Software Tools	[52], [53]	Forensics	No	Interpretation, TracerFire, 26 participants, 2 CSXs
Number Returns to a Previous Software Tool	[52], [53]	Forensics	No	Interpretation, TracerFire, 26 participants, 2 CSXs
Number of valid flags submitted	[19]	Basic encryption, access control, protocol analysis, web security, RE	Yes	Correlations over 3 datasets to final grade CTF, 3 iterations, no details of flags provided
Total number of logins over two months per weekday	[20]	Various cybersecurity knowledge	Partial	Interpretation, feedback survey CTF with lectures, no details of flags provided
Total number of logins over two months per hour	[20]	Various cybersecurity knowledge	Partial	Interpretation, feedback survey CTF with lectures, no details of flags provided
Tools, Commands and Methods Used by Learner				
Commandline (nmap, Linux bash history)	[15]	Network reconnaissance	No	Interpretation, 24 teams of students, 2 classes 2 schools
Reconnaissance Commands	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
Use of Internet browsers	[16]	Forensics	No	Interpretation, TracerFire, 11 participants
Frequency of software tools use	[16]	Forensics	No	Interpretation, TracerFire, 11 participants
Number of software tools used	[16]	Forensics	No	Interpretation, TracerFire, 11 participants
Type of software (general vs specialised) tools used	[16]	Forensics	No	Interpretation, TracerFire, 11 participants
Choice of tools used	[56]	Efficiency of student actions	No	Idea proposed, no measurements
Programming languages used	[56]	Efficiency of student actions	No	Idea proposed, no measurements
Input logs				
Direct input (logs)	[56]	Not specified	No	Idea proposed, no measurements
String similarity metrics (using e.g., Levenshtein distance)	[56]	Efficiency of student actions	No	Idea proposed, no measurements
Blocks of activity	[52], [53]	Forensics	No	Interpretation, TracerFire, 26 participants, 2 exercises
Log data: frequent itemsets to learning behaviors	[22]	Various cybersecurity knowledge	Yes	Data analysis: 7-hours competition collecting 8,257 logs from CTF Server and 407,623 logs from GRR Server
Log data: start/end, incorrect flags, hints used, skipping level, displaying solution, game ID	[24]	Various cybersecurity skills	No	No detailed metrics analysis
Network data: dst_ip, dst_port, ip_proto, ip_len, signature, signature_gen, priority, class, status	[50]	IDS alerts, network sessions, or top destination IP addresses	No	Nave Bayes and decision tree algorithms. For results validation, k-fold cross validation, 10 iterations, 17 students, 1 lab
Service status (active/non-active, vulnerable/not-vulnerable)	[38]	Various cybersecurity skills	No	Idea proposed, no measurements
Network traffic, teams attacked, vulnerable services, teams activity rate and strategy used, network traffic peaks, protocols usage, etc.	[38]	Various cybersecurity skills	No	Recommendation to look at logs to statistically evaluate what attacks were most efficient, possible damage caused, threats easily defended and a team's performance to prior CSX
Tags: OS, programming language, vulnerability, language, associated CVE, tools	[23]	Various cybersecurity knowledge	No	Discussion and experience
Logs: automatic scoring, pcap, chats/emails screen capture, video and audio	[81]	Various cybersecurity skills	Yes	Statistical analysis (team performance)
Joint Information Exchange Environment and Chat logs (hashtags)	[80]	Defending network against attack	No	Bivariate regression analysis, multivariate regression analysis, and principal component analysis
Other—rankings, indexes, indicators, manual				
Success rate (correct answer) from total challenges	[16]	Forensics	No	Interpretation, TracerFire, 11 participants
Abandonment rate of challenges	[16]	Forensics	No	Interpretation, TracerFire, 11 participants
Challenge submission accuracy	[16]	Forensics	No	Interpretation, TracerFire, 11 participants
Submission data and completion rate of challenges	[21]	Forensics, cryptography, RE, web and scripting exploitation, binary exploitation	No	Discussion and experience, PicoCTFs, 1588 participating teams, survey data
Total Similarity (based on several similarity indexes)	[49]	Defending network against attack	No	Interpretation, 4 participants, self-developed cyber range
Proxy indicators (e.g., observable indicators of adversarial thinking)	[20]	Various cybersecurity knowledge	No	Recommendation, overall CTF evaluation model
Scalar unit for each mitigation completed under cooperation, algorithm	[36]	Attack mitigation	No	Algorithm to measure skill improvement, SO tool
Soft Skills / Cognitive Metrics				
Teamwork: task ownership changes, dashboard, timecounter, option to mark challenge as difficult or solved	[23]	Various cybersecurity knowledge, teamwork	No	Discussion and experience
Awareness measure/critical thinking/decision making	[70], [43], [72], [73]	Understanding threats and systems, making desicions	No	Discussion and experience
Cognitive Agility Index	[82]	Individual cognitive performance	Yes	Regression analyses, science based, validation of 31 participants

and big datasets. The large amount of data generated by automatic logs and sensors necessitates efficient and automated data and LA techniques. There may not be enough traces to identify learning patterns (e.g., short time of detection, gaps in time-line) and data may be very diverse (e.g., different OS, applications). Therefore, identification of the relevant learning traces requires techniques that can deal with such imbalance and diversity. To combine multiple datasets and formats, e.g., technical and timing data with self-reported learner data no detailed descriptions or methods are provided how data is processed and time-synced. However, some examples were found about aligning timestamps of datasets, such as [53], [14], [22].

Also, the CSXs and related studies mostly work over a short period but it is known that short-term interventions are not particularly effective at affecting behavioral change [88]. Thus longitudinal studies are needed to evidence learning and behavior change as result of the exercises, and also to separate from other learning.

6. Conclusion

The opportunity to improve the learning in CSXs as part of educational effort is missed without considering the learners experience, different learning styles and pace, and the impact of the learning environment. The application of learning analytics and analysing digital datasets can provide a deeper understanding of learning behaviour and lead to evidence-based improvement. The consideration of LA aspects is also vital for the cyber range developers, as they design the technological foundation of instrumenting exercises that enable the application effective LA methods.

We proposed a LA reference model to assist in implementing LA into the CSXs life-cycle to achieve a more adaptive design and measurement using evidence-based data from the learning environment. As a practical starting point, we shared extensive related work overview of existing research describing some aspects of learning analytics process and the analysis of empirical evidence from the digital datasets to assist in implementing the model across all exercise types. We described the learning indicators (metrics) used for evidencing learning in CSXs, with focus on analytical evidence from digital dataset. Such metrics are mainly simple technical measures (time, number of attacks mitigated, availability of service, etc.) that are not necessarily validated and may not evidence effective learning (i.e., metacognition achieved). With LA and evidence-based measurement, we also need keep in mind and validate that what we measure (i.e., metrics used) actually help learners to learn. In turn, the validated metrics have the potential to provide more detailed and evidence-based input that form an integral part of the comprehensive training evaluations.

Further work should seek to identify and validate what learning metrics are evidencing the learning process and learning improvement in CSXs. Understanding the current use of learning analytics in CSXs is expected to help setting the baseline for further research and practical implementation by combining two evolving disciplines. By doing this, the cybersecurity community can establish more evidence-based and systematic approach for the evaluation of learning impact that will enable the design of more effective learning experiences.

Acknowledgments. This work was partially supported by the ECHO project which has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement no 830943.

References

[1] Nita G Brooks, Timothy H Greer, and Steven A Morris, "Information systems security job advertisement analysis: Skills review and implications for information systems curriculum," *Journal of Education for Business*, vol. 93, no. 5, pp. 213–221, 2018.

[2] George Siemens, "Learning analytics: The emergence of a discipline," *American Behavioral Scientist*, vol. 57, no. 10, pp. 1380–1400, 2013.

[3] Kirsty Kitto, Simon Buckingham Shum, and Andrew Gibson, "Embracing imperfection in learning analytics," in *Proceedings of the 8th International Conference on LAK*. ACM, 2018, pp. 451–460.

[4] Zacharoula Papamitsiou and Anastasios A Economides, "Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence.," *Journal of Educational Technology & Society*, 2014.

[5] Krzysztof Cabaj, Dulce Domingos, Zbigniew Kotulski, and Ana Respício, "Cybersecurity education: Evolution of the discipline and analysis of master programs," *Computers & Security*, vol. 75, pp. 24–35, 2018.

[6] Adrien Ogee, Razvan Gavrila, Panagiotis Trimintzios, Vangelis Stavropoulos, and Alexandros Zacharis, "The 2015 report on national and international cyber security exercises," ENISA, https://www.enisa.europa.eu/publications/latest-report-on-national-and-international-cyber-security-exercises.

[7] Art Conklin, "Cyber defense competitions and information security education: An active learning solution for a capstone course," in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences*. IEEE, 2006, vol. 9, pp. 220b–220b.

[8] Victor-Valeriu Patriciu and Adrian Constantin Furtuna, "Guide for designing cyber security exercises," in *Proceedings of the 8th WSEAS International Conference on E-Activities and information security and privacy*, 2009, pp. 172–177.

[9] Gordon Wells and Guy Claxton, *Learning for life in the 21st century: Sociocultural perspectives on the future of education*, John Wiley & Sons, 2008.

[10] Hanne Foss Hansen, "Choosing evaluation models: a discussion on evaluation design," *Evaluation*, vol. 11, no. 4, pp. 447–462, 2005.

[11] Jason Kick, "Cyber exercise playbook," 2014, MITRE Corporation, https://www.mitre.org/publications/technical-papers/cyber-exercise-playbook.

[12] Mika Karjalainen, Tero Kokkonen, and Samir Puuska, "Pedagogical aspects of cyber security exercises," in *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2019, pp. 103–108.

[13] Mohamed Amine Chatti, Anna Lea Dyckhoff, Ulrik Schroeder, and Hendrik Thüs, "A reference model for learning analytics," *International Journal of Technology Enhanced Learning*, vol. 4, no. 5-6, pp. 318–331, 2013.

[14] Kaie Maennel, Rain Ottis, and Olaf Maennel, "Improving and measuring learning effectiveness at cyber defense exercises," in *Nordic Conference on Secure IT Systems*. Springer, 2017, pp. 123–138.

[15] Richard Weiss, Michael E Locasto, and Jens Mache, "A reflective approach to assessing student performance in cybersecurity exercises," in *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*. ACM, 2016, pp. 597–602.

[16] Austin Silva, Jonathan McClain, Theodore Reed, Benjamin Anderson, Kevin Nauer, Robert Abbott, and Chris Forsythe, "Factors impacting performance in competitive cyber exercises," in *Proceedings of the Interservice/Interagency Training, Simulation and Education Conference, Orlando, FL*, 2014.

[17] Valdemar Švábenskỳ, Jan Vykopal, and Pavel Čeleda, "Toward an automated feedback system in educational cybersecurity games," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE19)*, 2019.

[18] Margus Ernits, Kaie Maennel, Sten Mäses, Toomas Lepik, and Olaf Maennel, "From simple scoring towards a meaningful interpretation of learning in cybersecurity exercises," in *ICCWS 2020 15th International Conference on Cyber Warfare and Security*, 2020.

[19] Tom Chothia and Chris Novakovic, "An offline capture the flag-style virtual machine and an assessment of its value for cybersecurity education," *USENIX Summit on Gaming, Games, and Gamification in Security Education*, 2015.

[20] Ronald S Cheung, Joseph Paul Cohen, Henry Z Lo, Fabio Elia, and Veronica Carrillo-Marquez, "Effectiveness of cybersecurity competitions," in *Proceedings of the International Conference on Security and Management*, 2012, p. 1.

[21] Peter Chapman, Jonathan Burket, and David Brumley, "Picoctf: A game-based computer security competition for high school students," in *3GSE*, 2014.

[22] Shian-Shyong Tseng, Sung-Chiang Lin, Ching-Hao Mao, Tsung-Ju Lee, Guan-Wei Qiu, and Ming-Han Lin, "An ontology guiding assessment framework for hacking competition," in *10th International Conference on Ubi-media Computing and Workshops*. IEEE, 2017, pp. 1–4.

[23] Nicholas Capalbo, Theodore Reed, and Michael Arpaia, "Rtfn: enabling cybersecurity education through a mobile capture the flag client," in *Proceedings of the International Conference on Security and Management*, 2011.

[24] Jan Vykopal and Milos Barták, "On the design of security games: From frustrating to engaging learning," in *ASE@USENIX Security Symposium*, 2016.

[25] Jan Vykopal, Martin Vizváry, Radek Oslejsek, Pavel Celeda, and Daniel Tovarnak, "Lessons learned from complex hands-on defence exercises in a cyber range," in *Frontiers in Education Conference*. IEEE, 2017, pp. 1–8.

[26] Valdemar Švábenskỳ, "Analyzing user interactions with cybersecurity games," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. ACM, 2019, pp. 1295–1295.

[27] Valdemar Švábenskỳ, Jan Vykopal, and Pavel Celeda, "Towards learning analytics in cybersecurity capture the flag games," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. ACM, 2019, pp. 1255–1255.

[28] Radek Ošlejšek, Vít Rusňák, Karolína Burská, Valdemar Švábenskỳ, and Jan Vykopal, "Visual feedback for players of multi-level capture the flag games: Field usability study," *arXiv preprint arXiv:1912.10781*, 2019.

[29] Elena Sitnikova, Ernest Foo, and Rayford B Vaughn, "The power of hands-on exercises in scada cyber security education," in *IFIP World Conference on Information Security Education*. Springer, 2009, pp. 83–94.

[30] Ryan Richards, Abdullah Konak, Michael R Bartolacci, and Mahdi Nasereddin, "Collaborative learning in virtual computer laboratory exercises," *Network, Security*, vol. 155, pp. 9, 2015.

[31] Pavel Čeleda, Jakub Čegan, Jan Vykopal, and Daniel Tovarňák, "Kypo–a platform for cyber defence exercises," *M&S Support to Operational Tasks Including War Gaming, Logistics, Cyber Defence. NATO Science and Technology Organization*, 2015.

[32] Radek Ošlejšek, Dalibor Toth, Zdenek Eichler, and Karolína Burská, "Towards a unified data storage and generic visualizations in cyber ranges," in *ECCWS 2017 16th European Conference on Cyber Warfare and Security*, 2017, p. 298.

[33] Radek Ošlejšek, Jan Vykopal, Karolína Burská, and Vít Rusňák, "Evaluation of cyber defense exercises using visual analytics process," in *Frontiers in Education Conference*. IEEE, 2018, pp. 1–9.

[34] Jan Vykopal, Radek Ošlejšek, Karolína Burská, and Kristína Zákopčanová, "Timely feedback in unstructured cybersecurity exercises," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. ACM, 2018, pp. 173–178.

[35] Michael Thompson and Cynthia Irvine, "Active learning with the cyberciege video game," in *Proceedings of the 4th conference on Cyber security experimentation and test*. USENIX Association, 2011, pp. 10–10.

[36] P. Legato and R. M. Mazza, "Modeling and simulation of cooperation and learning in cyber security defense teams," in *Proceedings - 31st European Conference on Modelling and Simulation, ECMS 2017*, 2017, pp. 502–509.

[37] Denise Nicholson, Lauren Massey, R O'Grady, and E Ortiz, "Tailored cybersecurity training in lvc environments," in *MODSIM World Conference, Virginia Beach, VA*, 2016.

[38] Andre LM Santos, Fred A Freitas, Leandro J Martins, Rodrigo L Magalhes, and Saulo RF Hachem, "Towards a cloud-based cyber war simulator," in *Proceedings of SBGames*, 2012.

[39] Angelo Furfaro, Antonio Piccolo, Andrea Parise, Luciano Argento, and Domenico Saccà, "A cloud-based platform for the emulation of complex cybersecurity scenarios," *Future Generation Computer Systems*, vol. 89, pp. 791–803, 2018.

[40] Affan Yasin, Lin Liu, Tong Li, Jianmin Wang, and Didar Zowghi, "Design and preliminary evaluation of a cyber security requirements education game (sreg)," *Information and Software Technology*, vol. 95, pp. 179–200, 2018.

[41] Natalie Coull, Iain Donald, Ian Ferguson, Eamonn Keane, Thomas Mitchell, Oliver V Smith, Erin Stevenson, and Paddy Tomkins, "The gamification of cybersecurity training," in *International Conference on Technologies for E-Learning and Digital Entertainment*. Springer, 2017, pp. 108–111.

[42] Kyle E Stewart, Jeffrey W Humphries, and Todd R Andel, "Developing a virtualization platform for courses in networking, systems administration and cyber security education," in *Proceedings of the 2009 Spring Simulation Multiconference*, 2009, p. 65.

[43] F. Wu and C. Gonzalez, "How could cyber analysts learn faster and make better decisions?," in *24th Conference on Behavior Representation in Modeling and Simulation, BRiMS 2015, co-located with the International Social Computing, Behavioral Modeling and Prediction Conference, SBP 2015*, 2015, pp. 35–42.

[44] Bin Zhang, Kamran Shafi, and Hussein A Abbass, "Robo-teacher: A computational simulation based educational system to improve cyber security," in *Robot Intelligence Technology and Applications 2012*. Springer, 2013.

[45] Agnė Brilingaitė, Linas Bukauskas, Virgilijus Krinickij, and Eduardas Kutka, "Environment for cybersecurity tabletop exercises," in *ECGBL 2017 11th European Conference on Game-Based Learning*, 2017, p. 47.

[46] Rain Ottis, "Light weight tabletop exercise for cybersecurity education," *Journal of Homeland Security and Emergency Management*, 2014.

[47] Irfan Ahmed and Vassil Roussev, "Peer instruction teaching methodology for cybersecurity education," *IEEE Security & Privacy*, vol. 16, no. 4, pp. 88–91, 2018.

[48] S. Morgan and B. Lagesse, "Dynamically generated virtual systems for cyber security education," in *Proceedings of the International Conference on Cloud Security Management*, 2015, vol. Jan, pp. 187–193.

[49] William Aubrey Labuschagne and Marthie Grobler, "Developing a capability to classify technical skill levels within a cyber range," in *ECCWS 2017 16th European Conference on Cyber Warfare and Security*, 2017, p. 224.

[50] Emin Caliskan, Unal Tatar, Hayretdin Bahsi, Rain Ottis, and Risto Vaarandi, "Capability detection and evaluation metrics for cyber security lab exercises," in *ICMLG2017 5th International Conference on Management Leadership and Governance*. Academic Conferences and publishing limited, 2017, p. 407.

[51] Kelly J Neville and Jeremiah T Folsom-Kovarik, "Recommendation across many learning systems to optimize teaching and training," in *International Conference on Applied Human Factors and Ergonomics*. Springer, 2018.

[52] Robert G Abbott, Jonathan McClain, Benjamin Anderson, Kevin Nauer, Austin Silva, and Chris Forsythe, "Log analysis of cyber security training exercises," *Procedia Manufacturing*, vol. 3, pp. 5088–5094, 2015.

[53] Robert G Abbott, Jonathan T McClain, Benjamin Robert Anderson, Kevin S Nauer, Austin Ray Silva, and James C Forsythe, "Automated performance assessment in cyber training exercises," Tech. Rep., Sandia National Laboratories, Albuquerque, NM, 2015.

[54] Adrian Furtună, Victor-Valeriu Patriciu, and Ion Bica, "A structured approach for implementing cyber security exercises," in *8th International Conference on Communications*. IEEE, 2010, pp. 415–418.

[55] Erik Moore, Steven Fulton, and Dan Likarish, "Evaluating a multi agency cyber security training program using pre-post event assessment and longitudinal analysis," in *IFIP World Conference on Information Security Education*. Springer, 2017, pp. 147–156.

[56] Sten Mäses, Bil Hallaq, and Olaf Maennel, "Obtaining better metrics for complex serious games within virtualised simulation environments," in *European Conference on Games Based Learning*, 2017, pp. 428–434.

[57] Nina Wilhelmson and Thomas Svensson, *Handbook for planning, running and evaluating information technology and cyber security exercises*, Försvarshögskolan (FHS), 2011.

[58] "ISO 22398:2013 Societal Security—Guidelines for Exercises," International Organization for Standardization, https://www.iso.org/standard/50294.html.

[59] Agnė Brilingaitė, Linas Bukauskas, and Aušrius Juozapavičius, "A framework for competence development and assessment in hybrid cybersecurity exercises," *Computers & Security*, p. 101607, 2019.

[60] Menelaos Katsantonis, Panayotis Fouliras, and Ioannis Mavridis, "Conceptual analysis of cyber security education based on live competitions," in *Global Engineering Education Conference*. IEEE, 2017, pp. 771–779.

[61] Melissa Dark and Jelena Mirkovic, "Evaluation theory and practice applied to cybersecurity education," *IEEE Security & Privacy*, vol. 13, no. 2, 2015.

[62] Jelena Mirkovic, Melissa Dark, Wenliang Du, Giovanni Vigna, and Tamara Denning, "Evaluating cybersecurity education interventions: Three case studies," *IEEE Security & Privacy*, vol. 13, no. 3, pp. 63–69, 2015.

[63] Julia Nevmerzhitskaya, Elisa Norvanto, and Csaba Virag, "High impact cybersecurity capacity building," in *The International Scientific Conference eLearning and Software for Education*. "Carol I" National Defence University, 2019, vol. 2, pp. 306–312.

[64] Erik L Moore, Steven P Fulton, Roberta A Mancuso, Tristen K Amador, and Daniel M Likarish, "A short-cycle framework approach to integrating psychometric feedback and data analytics to rapid cyber defense," in *IFIP World Conference on Information Security Education*. Springer, 2019, pp. 45–58.

[65] Petteri Taitto, Julia Nevmerzhitskaya, and Csaba Virag, "Using holistic approach to developing cybersecurity simulation environments," *eLearning & Software for Education*, vol. 4, 2018.

[66] Chris Hart, *Doing a Literature Review: Releasing the Research Imagination*, Sage, 2018.

[67] Chris Eagle, "Computer security competitions: Expanding educational outcomes," *IEEE Security & Privacy*, vol. 11, no. 4, pp. 69–71, 2013.

[68] Brandon Mauer, William Stackpole, and Daryl Johnson, "Developing small team-based cyber security exercises," in *The International Conference on Security and Management, Las Vegas*, 2012.

[69] Jin-Ning Tioh, Mani Mina, and Douglas W Jacobson, "Cyber security training a survey of serious games in cyber security," in *Frontiers in Education Conference*. IEEE, 2017, pp. 1–5.

[70] John Burris, Wesley Deneke, and Brandon Maulding, "Activity simulation for experiential learning in cybersecurity workforce development," in *International Conference on HCI in Business, Government, and Organizations*. Springer, 2018, pp. 17–25.

[71] Erjon Zoto, Stewart Kowalski, Christopher Frantz, Edgar Lopez-Rojas, and Basel Katt, "A pilot study in cyber security education using cyberaims: A simulation-based experiment," in *IFIP World Conference on Information Security Education*. Springer, 2018, pp. 40–54.

[72] Edgar A. Lopez-Rojas Mazaher Kianpour Erjon Zoto, Stewart Kowalski, "Using a socio-technical systems approach to design and support systems thinking in cyber security education," in *CEUR Workshop Proceedings*, 2018, vol. 2107, pp. 123–128.

[73] Johan de Heer and Paul Porskamp, "Human behavior analytics from microworlds: the cyber security game," in *International Conference on Applied Human Factors and Ergonomics*. Springer, 2017, pp. 173–184.

[74] Carlos Arturo Martinez Forero, "Tabletop exercise for cybersecurity educational training; theoretical grounding and development," M.S. thesis, University of Tartu, 2016.

[75] Keith Cozine, "Thinking interestingly: the use of game play to enhance learning and facilitate critical thinking within a homeland security curriculum," *British Journal of Educational Studies*, vol. 63, no. 3, 2015.

[76] Emin Caliskan, M Oguzhan Topgul, and Rain Ottis, "Cyber security exercises: A comparison of participant evaluation metrics and scoring systems," *Strategic Cyber Defense: A Multidisciplinary Perspective*, vol. 48, 2017.

[77] Valdemar Švábenskỳ, Jan Vykopal, and Pavel Čeleda, "What are cybersecurity education papers about? a systematic literature review of sigcse and iticse conferences," in *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 2020, pp. 2–8.

[78] Shane Dawson and George Siemens, "Analytics to literacies: The development of a learning analytics framework for multiliteracies assessment," *The International Review of Research in Open and Distributed Learning*, 2014.

[79] Ioana Jivet, Maren Scheffel, Marcus Specht, and Hendrik Drachsler, "License to evaluate: Preparing learning analytics dashboards for educational practice," in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. ACM, 2018, pp. 31–40.

[80] Diane S Henshel, Gary M Deckard, Brad Lufkin, Norbou Buchler, Blaine Hoffman, Prashanth Rajivan, and Steve Collman, "Predicting proficiency in cyber defense team exercises," in *Military Communications Conference, MILCOM 2016-2016 IEEE*. IEEE, 2016, pp. 776–781.

[81] Dennis Andersson, Magdalena Granåsen, Thomas Sundmark, Hannes Holm, and Jonas Hallberg, "Exploratory sequential data analysis of a cyber defence exercise," in *Proceedings of the International Defense and Homeland Security Simulation Workshop*, 2011.

[82] Benjamin. Knox, Ricardo Lugo, Kirsi Helkala, Stefan Sütterlin, and Øyvind Jøsok, "Education for cognitive agility: Improved understanding and governance of cyberpower," in *European Conference on Information Warfare and Security, ECCWS*, 2018, vol. 2018-June, pp. 541–550.

[83] Travis Scheponik, Alan T Sherman, David DeLatte, Dhananjay Phatak, Linda Oliva, Julia Thompson, and Geoffrey L Herman, "How students reason about cybersecurity concepts," in *Frontiers in Education Conference*. IEEE, 2016, pp. 1–5.

[84] Vitomir Kovanović, Srećko Joksimović, Zak Waters, Dragan Gašević, Kirsty Kitto, Marek Hatala, and George Siemens, "Towards automated content analysis of discussion transcripts: A cognitive presence case," in *Proceedings of the 6th International Conference on LAK*. ACM, 2016, pp. 15–24.

[85] Kirsty Kitto, Mandy Lupton, Kate Davis, and Zak Waters, "Designing for student-facing learning analytics," *Australasian Journal of Educational Technology*, vol. 33, no. 5, pp. 152–168, 2017.

[86] Ran Liu and Kenneth R Koedinger, "Closing the loop: Automated data-driven cognitive model discoveries lead to improved instruction and learning gains.," *Journal of Educational Data Mining*, vol. 9, no. 1, 2017.

[87] Rakesh Verma, Murat Kantarcioglu, David Marchette, Ernst Leiss, and Thamar Solorio, "Security analytics: essential data analytics knowledge for cybersecurity professionals and students," *IEEE Security & Privacy*, vol. 6, pp. 60–65, 2015.

[88] Maurice Hendrix, Ali Al-Sherbaz, and Bloom Victoria, "Game based cyber security training: are serious games suitable for cyber security training?," *International Journal of Serious Games*, vol. 3, no. 1, pp. 53–61, 2016.