arXiv:2307.08582v1 [cs.CY] 13 Jul 2023

Applications of Educational Data Mining and Learning Analytics on Data From Cybersecurity Training

Valdemar Švábenský[1,2*], Jan Vykopal[1], Pavel Čeleda[1] and Lydia Kraus[1]

[1*]Institute of Computer Science, Masaryk University, Šumavská 15, Brno, 60200, Czech Republic.
[2*]Faculty of Informatics, Masaryk University, Botanická 68a, Brno, 60200, Czech Republic.

*Corresponding author(s). E-mail(s): svabensky@ics.muni.cz;
Contributing authors: vykopal@ics.muni.cz; celeda@ics.muni.cz; kraus@ics.muni.cz;

Abstract

Cybersecurity professionals need hands-on training to prepare for managing the current advanced cyber threats. To practice cybersecurity skills, training participants use numerous software tools in computer-supported interactive learning environments to perform offensive or defensive actions. The interaction involves typing commands, communicating over the network, and engaging with the training environment. The training artifacts (data resulting from this interaction) can be highly beneficial in educational research. For example, in cybersecurity education, they provide insights into the trainees' learning processes and support effective learning interventions. However, this research area is not yet well-understood. Therefore, this paper surveys publications that enhance cybersecurity education by leveraging trainee-generated data from interactive learning environments. We identified and examined 3021 papers, ultimately selecting 35 articles for a detailed review. First, we investigated which data are employed in which areas of cybersecurity training, how, and why. Second, we examined the applications and impact of research in this area, and third, we explored the community of researchers. Our contribution is a systematic literature review of relevant papers and their categorization according to the collected data, analysis

methods, and application contexts. These results provide researchers, developers, and educators with an original perspective on this emerging topic. To motivate further research, we identify trends and gaps, propose ideas for future work, and present practical recommendations. Overall, this paper provides in-depth insight into the recently growing research on collecting and analyzing data from hands-on training in security contexts.

Keywords: cybersecurity education, hands-on training, data science, literature survey, systematic literature review

1 Introduction

Cybersecurity education occurs at universities, extracurricular events, in the industry, and beyond. In all these cases, cybersecurity training is fundamentally practical. It involves exercising one's skills in applied computing topics, such as administration of operating systems, network attacks and defense, and secure programming. At the same time, cybersecurity is a complex, ever-evolving domain. Instructors and students need to keep up with the latest cyber threat landscape development through hands-on experience.

Because of its hands-on nature, cybersecurity training relies on interactive learning environments and testbeds: technologies that enable students to practice their skills in realistic computer systems. This opens opportunities for collecting objective evidence about learning processes, such as the used commands, system logs, and captured network traffic.

These pieces of evidence collected from cybersecurity training, also called the *training artifacts*, allow researchers to authentically reconstruct the actions that students performed while solving the training tasks. This provides the basis for achieving important educational goals, such as to:

- understand students' learning processes and approaches to solving the training assignments;
- assess students and measure their learning; and
- provide personalized, targeted instruction and feedback.

These goals align with the objectives of *educational data mining* (EDM) (Romero, Ventura, Pechenizkiy, & Baker, 2010) and *learning analytics* (LA) (Lang, Siemens, Wise, & Gašević, 2017), two growing disciplines that leverage data from educational contexts. They aim to better understand and improve teaching and learning (Hundhausen, Olivares, & Carter, 2017) by employing computing methods, such as:

- mathematical modeling,
- automated data analysis,
- mining of patterns and processes,
- natural language processing, and
- machine learning.

EDM/LA research findings support more effective training of security professionals, who are direly needed to handle the current cyber threats.

Since educational research is rooted in social sciences (Malmi et al., 2010), EDM/LA researchers traditionally collected and analyzed data from questionnaires and interviews (Lang et al., 2017; Romero et al., 2010). However, these data may not always constitute objective evidence. Students can, intentionally or unintentionally, report behavior or attitudes that differ from the truth. Therefore, it is difficult to ensure the validity and reliability of conclusions drawn from these data (Romero et al., 2010, Chapter 8.2.1). That is why we focus on training artifacts measured in interactive learning environments and their application in cybersecurity education research.

1.1 Goal of This Paper

Our goal is to understand the current landscape of EDM/LA in hands-on cybersecurity education. We seek to provide an original overview of this emerging research area that integrates technology and learning. To achieve this goal, we examine the published research on leveraging data from computer systems for cybersecurity training.

Specifically, we perform a systematic literature review (SLR) of research papers that analyze artifacts generated as a product of cybersecurity training. These topics have been recently gaining interest, but they were studied mostly in isolation. We contextualize and organize the research efforts and propose practical implications for further research.

Our findings are relevant in the cybersecurity domain, as well as related areas such as operating systems and networking. They may inspire other researchers who use training environments and collect data from them. Based on our SLR, researchers can learn what data sources to analyze and which approaches were covered in previous work.

1.2 Research Topics

We aim to examine three research topics in the surveyed papers.

1. *Domain and Data.* Since high-quality data are the key to conducting EDM/LA research, we focus on how the researchers addressed the issues of data collection and analysis. First, we explore the cybersecurity areas in which EDM/LA was applied. Then, we investigate how the data were collected, from which computer systems, and categorize the data into distinct types. We also look at the sample size, the time span of the data, how privacy was addressed, and methods for data analysis.

2. *Research Impact.* Next, we look at whether the methods or findings of the published research were applied in teaching practice, what were the main contributions, and what was the citation impact of the research.

3. *Research Community.* Finally, we examine the background of researchers in the field, summarize their chosen publication venues, and show how the selected papers are interrelated. Only highlights from this topic are

addressed, since this SLR does not primarily focus on bibliometric aspects (such as in (Zurita et al., 2020)).

Section 3.1 defines specific research questions within each of these topics. Answers to these questions provide an overview of this novel area for both new and well-established researchers.

1.3 Paper Structure

Section 2 explains the terminology and describes related primary and secondary studies. Section 3 details the methods for conducting the SLR and the review protocol. Section 4 presents and discusses the findings. Section 5 proposes ideas for future research and provides guidelines for EDM/LA researchers. Finally, Section 6 concludes and summarizes our contributions.

2 Background and Related Work

This section provides a brief background to cybersecurity training and EDM/LA. It presents related publications and compares them to this paper to explain how we differ from state of the art. Section 2.1 explains the popular formats of cybersecurity training to familiarize readers with the terminology. The overview of the related literature surveys focuses on two domains: *cybersecurity training* (Section 2.2) and *educational data analysis* (Section 2.3), since this paper is situated in their intersection.

2.1 Glossary of Hands-on Cybersecurity Training

In this paper, we consider all hands-on learning sessions during which the students practice their cybersecurity skills. The nomenclature for these sessions varies widely in the literature: they can be called labs, exercises, assignments, or practicals, and they can involve individual or team learning. Below, we particularly introduce two specific types of cybersecurity training: *Capture the Flag* (CTF) (Taylor, Arias, Klopchic, Matarazzo, & Dube, 2017) and *Cyber Defense Exercises* (CDXs) (Vykopal, Vizvary, Oslejsek, Celeda, & Tovarnak, 2017).

Gamification is popular in many application contexts, including education (Graham et al., 2020; Kasurinen & Knutas, 2018), and CTF is a flagship example of gamifying cybersecurity training. In a CTF game, the trainees solve cybersecurity assignments that yield flags: textual strings that are worth points. There are two main variations of the CTF format: *jeopardy* and *attack-defense* (Švábenský, Čeleda, Vykopal, & Brišáková, 2020).

In a jeopardy CTF, trainees choose the tasks from categories such as cryptography, reverse engineering, or forensics. They solve the tasks locally at their computers or interact with a remote server or network. This popular format is hosted even by tech giants like Google in their Google CTF (Google, 2021).

In an attack-defense CTF, teams of trainees each maintain an identical instance of a vulnerable computer system. Each team must protect its system

while exploiting vulnerabilities in the systems of other teams. Examples of these events include DEF CON CTF (DEF CON, 2021) and iCTF (Vigna et al., 2014).

CDXs such as Locked Shields (The NATO Cooperative Cyber Defence Centre of Excellence, 2021b), Crossed Swords (The NATO Cooperative Cyber Defence Centre of Excellence, 2021a), and Cyber Storm (Cybersecurity & Infrastructure Security Agency, 2018) aim at professionals, often from military or government agencies or dedicated cybersecurity teams. The trainees form teams whose roles are denoted by colors. *Blue* teams are responsible for maintaining and defending a complex network infrastructure against the attacks of a *red* team. Blue teams must preserve the availability of the network services for end-users. Both CTF and CDXs employ interactive learning environments that allow collecting vast arrays of valuable data, which we explore in this paper.

2.2 Literature Surveys in Cybersecurity Education

The closest paper to ours is a survey by Maennel (Maennel, 2020), who reviewed various data sources that can serve as evidence of learning in cybersecurity exercises. These data sources include timing information, command-line data, counts of events, and input logs. We chose a different methodology (see Section 3) and posed additional research questions to examine the current literature. Therefore, we provide a complementary and extended perspective.

Švábenský et al. (Švábenský, Vykopal, & Čeleda, 2020) performed a SLR of 71 cybersecurity education papers published at ACM SIGCSE and ACM ITiCSE conferences since 2010. They investigated which cybersecurity topics were published at these conferences, the teaching context, research methods, citations, and authors within the conferences' community. They found that the examined research primarily employed data from questionnaires and tests to evaluate student perceptions or learning gains. The difference is that this SLR focuses on the applications of EDM/LA and not on any specific venue or time period.

While the review (Švábenský, Vykopal, & Čeleda, 2020) focused mostly on university education, a review by Khando et al. (Khando, Gao, Islam, & Salman, 2021) focused on security awareness in organizations. They discovered that various methods, such as gamification and theoretical models, are used to enhance the security awareness of employees. Yet, the paper did not examine the data sources that can be mined in these security awareness programs.

Yamin et al. (Yamin, Katt, & Gkioulos, 2020) surveyed cyber ranges and security testbeds, platforms that provide technical infrastructure for cybersecurity training. They found that most of these platforms use various data collection mechanisms, including event logging and network layer monitoring. Kucek and Leitner (Kucek & Leitner, 2020), on the other hand, compared the functionality of open-source environments for conducting CTF sessions. Again, from the data collection perspective, most environments log statistics such as

the number of solved challenges and scoring data. However, none of the related papers in this section addressed the research questions we pose in Section 3.1.

2.3 Literature Surveys in EDM/LA and Computing Education

Several literature reviews about EDM/LA were published in the past few years (Liñán & Pérez, 2015). The covered topics include evaluation of LA interventions (Knobbout & Van Der Stappen, 2020), LA-driven learning design (Mangaroska & Giannakos, 2019), and LA dashboards (Matcha, Uzir, Gašević, & Pardo, 2020). A survey of 240 EDM works (Peña-Ayala, 2014) identified the most prominent approaches used in EDM research, which include Bayes theorem, decision trees, instances-based learning, and hidden Markov models. There is also a SLR of methods, benefits, and challenges of LA (Nunn, Avella, Kanai, & Kebritchi, 2016).

A paper related to ours is a thorough literature review of EDM/LA in programming (Ihantola et al., 2015). It evaluated the content and quality of 76 papers, examining the information that "can be gained through the analysis of programming data", and "which of that data can be collected and analyzed automatically". These data include keystrokes, line edits, program compilation, program execution, and more. Our paper also focuses on data collection and analysis, however, in cybersecurity training. Moreover, we evaluate the impact and applications of the published research and examine the research community.

Another thorough survey is by Luxton-Reilly et al. (Luxton-Reilly et al., 2018), who reviewed and classified 1666 publications on introductory programming education. The survey highlighted that programming data are used to examine students' compilation behavior, code correctness, and code style. These aspects are studied to identify student competencies and difficulties, predict their performance, and recognize demotivation, among other use cases.

Margulieux et al. (Margulieux, Ketenci, & Decker, 2019) reviewed 197 texts to identify variables measured in computing education papers. These include student performance and information about the timing, progress, and collaboration, for example.

Lastly, Papamitsiou et al. (Papamitsiou, Giannakos, Simon, & Luxton-Reilly, 2020) analyzed keywords in 1274 computing education papers to discover clusters of recurring topics. Among the most frequent are assessment, introductory programming, games, and computational thinking.

All these papers indicate a vast potential for EDM/LA in computing education, yet little is known about its application in the field of cybersecurity training. Our article aims to close this gap by examining this emerging topic.

3 Method of Conducting the Systematic Literature Review

To perform this study, we followed the well-established guidelines for conducting a SLR (Kitchenham & Charters, 2007; Moher, Liberati, Tetzlaff, Altman, & Group, 2009). We also consulted recommendations for a systematic mapping study (Petersen, Feldt, Mujtaba, & Mattsson, 2008; Petersen, Vakkalanka, & Kuzniarz, 2015) and a literature review section for a Ph.D. dissertation (Randolph, 2009). This section presents the SLR protocol, which specifies the research questions, search process, and criteria for including the discovered papers.

3.1 Research Questions

We seek to answer the following research questions to understand the state of the art at the intersection of EDM/LA and cybersecurity training.

Research Topic 1: Domain and Data

RQ1.1 In which *areas* of cybersecurity training was EDM/LA applied?

RQ1.2 What was the *intent* of the data collection?

RQ1.3 From which computer systems or *environments* were the data collected?

RQ1.4 What *types of data* were collected from these systems?

RQ1.5 From *how many students* were the data collected?

RQ1.6 What was the *time span* of the data? In other words, how long did the educational activity last while the data were collected?

RQ1.7 Since EDM/LA involves collecting data about people, did the research address data anonymization and *privacy* preservation?

RQ1.8 Which *analysis methods* were applied to the collected data?

Research Topic 2: Research Impact

RQ2.1 In which *educational context* was the research practically applied?

RQ2.2 What were the *contributions* of the research?

RQ2.3 What were the *supplementary materials* of the research?

RQ2.4 How much was the research *cited*?

Research Topic 3: Research Community

RQ3.1 Who were the *authors* of the research, and what were their affiliations?

RQ3.2 What are the characteristics of the *conferences and journals* they choose for publishing?

RQ3.3 How much did the members of the community *cite* each other?

3.2 Identifying Sources for the Automated Search for Papers

We decided not to search for papers in the databases of individual publishers, such as the ACM Digital Library or IEEE Xplore, to avoid inaccuracies and conflicts when merging the results. Instead, we considered three aggregate databases: Web of Science, Scopus, and Google Scholar.

We ultimately used Scopus (Elsevier, 2021), since it indexes a representative portion of the databases of individual publishers. We did not choose Web of Science because it does not index several years of relevant educational conferences, such as ACM SIGCSE. We also omitted Google Scholar since it indexes many lower-quality publications, such as non-peer-reviewed papers.

3.3 Selecting the Keywords for the Automated Search

When defining the search terms, we aimed to cover the intersection of cybersecurity education and data analysis. We collected the keywords from multiple sources: previously known relevant papers, our expertise, and the knowledge of three cybersecurity experts independent from the paper authors. After multiple iterations and test searches, we established the search query in Figure 1.

```
(
  (
    (cybersecurity OR "cyber security" OR "computer security"
      OR "information security" OR "network security")
    AND
    (educat* OR teach* OR instruct* OR student* OR learner OR exercis*)
  )
  OR
  ("capture the flag" OR "cyber defense exercise" OR "cyber defence
    exercise" OR "security training" OR "security exercise" OR "cyber range")
)
AND
(analy* OR evaluat* OR examin*)
```

Fig. 1 The query for the automated search for papers in the Scopus database. Asterisks represent wildcards, and the search is case-insensitive.

After several pilot searches, we excluded the keywords `learn*` and `train*`, as they matched hundreds of general machine learning papers about deep learning or training classifiers. It is important to note that this exclusion did not eliminate educational papers. Publications about teaching or learning included at least one of the other educational keywords we used, such as `educat*`, `teach*`, or `student*`. Moreover, we added keywords specific to cybersecurity education, such as `security training` or `security exercise`.

We also excluded the keyword `security`, since it yielded too many irrelevant papers (for example, about fire safety or physical security). Finally, we

removed keywords related only to operating systems or networking, since during test searches, the candidate set of results was huge. Nevertheless, the query remained broad enough to avoid the risk of missing a relevant paper.

3.4 Performing the Automated Search for Candidate Papers

Figure 2 shows an overview of the SLR process. We started by submitting the query in Figure 1 to the online database Scopus (Elsevier, 2021). We restricted the search to titles, abstracts, and keywords of papers in conference proceedings or journals, in the area of computer science or engineering, and in the English language. Then, we exported the results as bibliographic records (in the `bib` format) to the Mendeley reference manager (Mendeley, 2021).
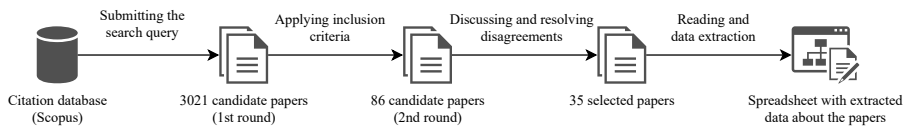
Fig. 2 Overview of the steps of the systematic literature review, along with the numbers of papers at each stage.

We performed the first search on March 29, 2019. Immediately afterward, we subscribed to Scopus e-mail notifications that informed us about newly indexed papers, which we gradually added to the candidate set. We stopped adding new candidates on September 24, 2021.

This process yielded 3021 candidate papers. To evaluate the search, we checked that the candidate set included relevant papers that we knew from our previous research in cybersecurity education.

3.5 Defining the Inclusion and Exclusion Criteria

After multiple iterations and pilot tests, we defined five selection criteria. We explain the rationale behind each criterion and provide examples of exclusion when necessary.

1. The paper must have *full text* available and be at least *four pages* long. Shorter papers lacked space to provide detailed information needed to answer our research questions.

2. The paper must *report on collected data* that originate from *human interaction* with a *computer system* for cybersecurity training. By a computer system, we mean a hardware, or virtualized, or cloud infrastructure, in which the student interacts with software applications. This comprises either *individual* or *team* learning that occurs during labs, training sessions, simulations, competitions, exercises, and the like.

   We excluded papers that dealt with video and board games. Although these games have educational potential, they do not emulate realistic

cybersecurity operations that involve cybersecurity tools and their monitoring. We also excluded purely system design papers that only proposed a data collection/analysis toolchain but did not apply it in practice, not even in author testing. The reasoning from our experience was that the system might theoretically look good as a proposal on paper, but its true capabilities can only be proven in practice. So, our research questions focus on practical demonstration and application.

3. The collected data must result either from interactions with an *operating system* and *applications* running in it (for example, recorded keystrokes, mouse clicks, memory dump, filesystem changes, or network traffic) or with some additional *training system* (such as timings of actions in a separate training interface).

   This means we excluded papers that reported only grades of students or only administered questionnaires (for example, (Chothia & Novakovic, 2015)). As we argued in Section 1, we consider these data often not representative of students' actions. For similar reasons, we also excluded papers focusing solely on the affective domain, such as emotion recognition (Imani & Montazer, 2019).

4. The data must be collected and analyzed *automatically* or at least semi-automatically. We excluded papers that involved only fully manual data processing, such as human graders observing the students and noting their actions (for example, (Rege et al., 2017)). This approach is time-consuming to replicate, does not scale, and is prone to errors. Therefore, our review targets computer technologies for automated data processing.

5. The data analysis must support an *educational goal*, for example, to assess students or help instructors understand the students' learning processes or behavior. This means we excluded papers that presented only performance testing of the learning environment.

3.6 Preliminary Reading and Applying the Inclusion Criteria

Two authors, independently of each other, preliminarily screened each of the 3021 candidate papers and applied the inclusion criteria. We followed the process in Figure 3.

3.7 Resolving Disagreements and Selecting Papers for Review

When the two readers finished, they compared their decisions. There were three possibilities:

- If both readers voted to include the paper, it was selected.
- If both readers voted to exclude the paper, it was rejected.
- If there was a disagreement, it was resolved by discussion. Afterward, the paper was either selected or rejected.

```
for each paper in the candidate set:
    read the title and abstract
    decide for inclusion or exclusion
    if decision cannot be made:
        read the introduction and conclusion
        decide for inclusion or exclusion
        if decision cannot be made:
            skim-read the rest of the paper
            decide for inclusion or exclusion
```

Fig. 3 The algorithm for selecting papers for the literature review from the candidate papers.

The readers initially agreed on 97.8% of papers (23 immediate selections and 2935 immediate rejections). We initially disagreed and further discussed the remaining 2.2% (63 papers). Our inter-rater agreement (Krippendorff, 2004) measured by Scott's π and Krippendorff's α (for nominal data) was 0.41, which is moderate. The coefficients were calculated using the Python NLTK module (Natural Language Toolkit (NLTK) Project, 2022).

3.8 Extracting Data from Selected Papers

We selected 35 papers for the detailed review. We read their full texts, extracted the information determined by our research questions, and recorded them in a spreadsheet. Section 4 presents the results.

4 Results and Discussion

Although the oldest candidate paper is from the year 1976, the oldest selected paper is from 2012. Moreover, more than two-thirds of the selected papers were published in 2017 or later. This implies that EDM/LA in cybersecurity is an arising topic that has recently started gaining traction, and it will likely continue in this growing trend.

We now answer the research questions from Section 3.1. Throughout this section, we refer to Table 1 and Table 2, which summarize the results.

Table 1 Overview of the goals of the 35 reviewed papers grouped by topics. The papers are identified by an arbitrary number based on the year of publication. For two papers Px and Py, if $x < y$, then Px was published before Py or in the same year.

Paper ID	Cybersecurity Topics (RQ1.1)	Goal of the Paper
P3 Chapman, Burket, and Brumley (2014)	offense, forensics	analyze the preferences, activity, and number of learners in a CTF
P5 Burket, Chapman, Becker, Ganas, and Brumley (2015)	offense	detect cheating in a CTF by analyzing sharing of solutions
P7 Weiss, Locasto, and Mache (2016)	offense	assess learners by visualizing their command history as a directed graph
P9 Vykopal and Barták (2016)	offense	determine what information can be predicted from logs of 260 trainees
P11 Tseng et al. (2017)	offense	analyze learners' behavior in a CTF to reveal their misconceptions
P12 Caliskan, Tatar, Bahsi, Ottis, and Vaarandi (2017)	offense	determine metrics from exercise logs that will predict students' grade
P14 Andreatos (2017)	offense	analyze students' network activity in a lab to review their actions
P16 Kont, Pihelgas, Maennel, Blumbergs, and Lepik (2017)	offense	provide and evaluate feedback for the attacking teams in a CDX
P18 Chothia, Holdcroft, Radu, and Thomas (2017)	offense	determine if storyline in a cybersecurity training improves learning
P19 Tian et al. (2018)	offense	provide trainees with situational awareness of the training
P21 Švábenský and Vykopal (2018a)	offense	determine if trainees fulfill prerequisites of security training
P22 Švábenský and Vykopal (2018b)	offense	analyze how trainees interact with security training tasks and tools
P25 Andreolini, Colacino, Colajanni, and Marchetti (2019)	offense	assess trainees by comparing their actions to a reference solution
P26 Falah, Pan, and Chen (2019)	offense	estimate the difficulty of attacks and measure skills of trainees
P28 Maennel, Mäses, Sütterlin, Ernits, and Maennel (2019)	offense, network security	assess students who apply to a cybersecurity master degree program
P31 Tobarra et al. (2020)	offense, forensics, network security	compare assessment of students who did / did not participate in a CTF
P34 Kaneko et al. (2020)	offense, forensics	evaluate an intensive cybersecurity course based on student performance
P35 Yett et al. (2020)	offense, secure programming	analyze how students collaborate in group programming tasks
P2 Reed, Nauer, and Silva (2013)	forensics	analyze score distribution, submission delay, and frustration in a CTF
P6 Abbott et al. (2015)	forensics	quantitatively analyze student actions and performance in security training
P1 Rupp et al. (2012)	network security	identify skill profiles of students based on logs and submitted commands
P23 Zeng, Deng, Hsiao, Huang, and Chung (2018)	network security	compare student grades with the time they spent working on lab tasks
P24 Deng, Lu, Chung, Huang, and Zeng (2018)	network security	adapt instruction to trainees' learning style, predict their performance
P27 Palmer (2019)	network security	automatically assess the quality of students' network configuration
P30 Sheng (2020)	network security	evaluate a custom machine learning model for assessing students
P33 Tobarra et al. (2020)	network security	evaluate how often and how long students interact with a training platform
P8 Granåsen and Andersson (2016)	incident response	assess performance, behavior, and progress of teams in a CDX
P10 Henshel et al. (2016)	incident response	determine proficiency metrics to assess performance of teams in a CDX
P15 Labuschagne and Grobler (2017)	incident response	assess trainees by comparing their command history with an ideal solution
P17 Maennel, Ottis, and Maennel (2017)	incident response	propose and apply a methodology for measuring learning in a CDX
P20 Kokkonen and Puuska (2018)	incident response	analyze communication patterns of defending teams in a CDX
P13 Weiss, Turbak, Mache, and Locasto (2017)	system administration	assess learners by visualizing their command history as a directed graph
P4 Nadeem, Allen, and Williams (2015)	secure programming	recommend reading to developers based on vulnerabilities in their code
P29 Espinha Gasiba, Lechner, and Pinto-Albuquerque (2020)	secure programming	compare two methods for measuring time to solve a challenge
P32 Almansoori et al. (2020)	secure programming	analyze how students and instructors use unsafe C/C++ functions

Table 2 Overview of the intent of data collection, collected data, analysis methods, and contributions of the 35 selected papers.

Paper ID	Intent (RQ1.2)	Collected Data (RQ1.4)	Analysis (RQ1.8)	Contribution (RQ2.2)
P1 Rupp et al. (2012)	assess	C, T	DS, ML	study of learning in a training platform
P2 Reed et al. (2013)	assess, inform	E, T	DS, NM	measure of situational awareness
P3 Chapman et al. (2014)	assess, inform	E, T	DS	open-source platform and challenges
P4 Nadeem et al. (2015)	support	C	DS, NM	architecture of a learning system
P5 Burket et al. (2015)	assess	E, T	DS, QA	open-source platform and challenges
P6 Abbott et al. (2015)	assess	H, A, N, E	DS	architecture of data collection infrastructure
P7 Weiss et al. (2016)	assess, support	C	QA	study of the utility of command history
P8 Granåsen and Andersson (2016)	assess	A, N, I, V	DS, NM	assessment model for defense exercises
P9 Vykopal and Barták (2016)	assess, inform	E, T	DS	study of the utility of game logs
P10 Henshel et al. (2016)	assess	N, I, T	DS, ML	assessment model
P11 Tseng et al. (2017)	assess	H	QA	study of students' behavior
P12 Caliskan et al. (2017)	assess	H, N	DS, ML	study of grading students
P13 Weiss et al. (2017)	assess, support	C	QA	study of the utility of command history
P14 Andreatos (2017)	assess	N	DS	study of monitoring student network activity
P15 Labuschagne and Grobler (2017)	assess	C, T	DS, NM	method for scoring
P16 Kont et al. (2017)	support	H, N	DS	framework for attackers' situational awareness
P17 Maennel et al. (2017)	assess	N, T	QA	method for measuring learning
P18 Chothia et al. (2017)	assess	C	DS	study of story improving engagement
P19 Tian et al. (2018)	assess, support	H, N, D, C	QA	study of the utility of command history
P20 Kokkonen and Puuska (2018)	assess	I, T	QA	tool for CDX organizers' situational awareness
P21 Švábenský and Vykopal (2018a)	assess	E, T	DS, ML, QA	study of predicting prerequisites
P22 Švábenský and Vykopal (2018b)	assess, inform	E, T	DS, QA	study of students' behavior
P23 Zeng et al. (2018)	assess	T	DS	study of factors that contribute to learning
P24 Deng et al. (2018)	assess, support	C, T	DS, ML	method for personalizing instruction
P25 Andreolini et al. (2019)	assess	H, A, N, C, T	DS, NM, QA	method for modeling and scoring training
P26 Falah et al. (2019)	assess	E, T	DS, PM	method for scoring / assessing performance
P27 Palmer (2019)	assess	D	DS	tool for assessing performance
P28 Maennel et al. (2019)	assess	E, T	QA	lessons learned from assessing students
P29 Espinha Gasiba et al. (2020)	assess, support, inform	E, T	DS, NM, PM, QA	methods for computing challenge solve time
P30 Sheng (2020)	assess	N	NM, ML	experimental comparison of two metrics
P31 Tobarra et al. (2020)	assess	E, T	DS	study of effect of CTF on grades
P32 Almansoori et al. (2020)	support	C	DS, NM	study of issues in C/C++ code at universities
P33 Tobarra et al. (2020)	inform	E, T	DS, IS	study of students' interactions with a platform
P34 Kaneko et al. (2020)	assess	H, A, V, T	QA	study of an exercise-based cybersecurity course
P35 Yett et al. (2020)	assess	E, T	DS, IS, PM, QA	study of students' approaches to collaborative tasks

Collected data: C = shell commands and program code, N = network logs and traces, H = host-based logs, A = application logs, D = disk and memory content, E = training events, I = interaction and communication, V = video, T = timestamps.
Analysis methods: DS = descriptive statistics, IS = inferential statistics, NM = numerical methods, ML = machine learning, PM = probabilistic modeling, QA = qualitative analysis.

4.1 Research Topic 1: Domain and Data

We start by looking at the first eight research questions about the data collection and analysis.

4.1.1 RQ1.1: Cybersecurity Topics

We categorize the papers into custom technical topics and also identify the topics from the CSEC2017 cybersecurity curriculum (Joint Task Force on Cybersecurity Education, 2017). The categorization intentionally omits soft skills such as critical thinking and teamwork, which are outside the scope of this SLR.

As Table 1 shows, 18 papers focus on teaching offensive security skills, including penetration testing, exploitation, network attacks, cryptographic attacks, and reverse engineering. 22 papers focused on defensive skills, which we divided into the following:

- *Network security* (P1, P23, P24, P27, P28, P30, P31, P33), which includes technical defensive skills, such as configuring networks, firewalls, and intrusion detection.
- *Incident response* (P8, P10, P15, P17, P20), which involves the network security skills applied while resolving a simulated cybersecurity incident.
- *Forensic analysis* and examining digital evidence (P2, P3, P6, P31, P34).
- *Secure programming* and preventing vulnerabilities (P4, P29, P32, P35).
- *System administration* (P13), which involves configuring a Linux system.

Next, we mapped the topics onto the Knowledge Areas of the CSEC2017 curriculum (Joint Task Force on Cybersecurity Education, 2017). The mapping revealed that *Connection security* and *System security* are dominantly represented (in 26 and 23 papers, respectively). Also present were *Data security* (7 papers), *Software security* (4 papers), and *Component security* (3 papers). *Human*, *Organizational*, and *Societal security* were not present due to our inclusion criteria. Interestingly, although programming topics are prevalent in computing education research, there were only four papers on secure programming in our dataset.

However, the topic mapping was sometimes difficult. Only a minority of papers stated the learning objectives or described the cybersecurity skills they aim to practice. For example, in P2, we were unsure about the content of the exercises. We assigned the paper in the Forensics category because it stated that "The challenges contained forensics data". Moreover, very few papers referenced a standardized cybersecurity curriculum, such as (CC2020 Task Force, 2020; Joint Task Force on Computing Curricula, Association for Computing Machinery (ACM) and IEEE Computer Society, 2013), when defining their learning objectives. Nevertheless, we managed to work with the information that was apparent from the paper text.

4.1.2 RQ1.2: Intent of Data Collection

We synthesized three main purposes of data collection:

- *Assess students*, that is, measure their performance, reveal their misconceptions, analyze their task solution patterns, or otherwise evaluate their actions. This was the goal of 31 papers, the vast majority.
- *Support learning* of students or provide feedback to them. This was the goal of 8 papers. Some publications had multiple goals, so 5 of these papers overlapped with the assessment papers.
- *Inform the training content creators* to provide them feedback about how the students approached the tasks. This was the goal of 6 papers.

Table 2 categorizes the selected papers based on the intent. Although several papers have the same overarching goal, they achieve it with different applications of EDM/LA. We gradually analyze various aspects of these applications in the following sections.

4.1.3 RQ1.3: Environments for Data Collection

A learning environment that allows automated data collection is a crucial prerequisite for any EDM/LA applications. Therefore, it was one of the aspects on which we focused in our review. Specifically, we observed four types of environments from which the data were collected:

- The *training infrastructure*, which is a physical or virtual environment that consists of one or more hosts with a standard operating system. The hosts are usually networked. This category includes cyber ranges (Yamin et al., 2020) and lab platforms, and 24 papers collected data from them.
- A *software application* that simulates a network environment. This was applicable only for P1 and P27.
- *Learning management system* (LMS), which is a web-based technology that facilitates the training, such as a CTF platform (Kucek & Leitner, 2020). LMS allows collecting the solutions to tasks submitted by students. This applied to 8 papers.
- *External sources* of data. This was applicable only for P4 and P32 that collected source code from repositories.

4.1.4 RQ1.4: Collected Data

The data collected from learning contexts were largely heterogeneous, demonstrating the diverse possibilities that EDM/LA offers. We synthesized nine categories denoted by capital letters used in Table 2 and Figure 4.

The following data were collected from the training infrastructure or simulation software as a result of the students' direct interaction with the training environment:

- *Shell commands and program code* (C), including that from external sources (10 papers).
- *Network logs and traces* (N), including packet captures and intrusion detection system logs (10 papers).
- *Host-based logs* (H), which include Syslog, audit logs, event logs, CPU and memory usage, and process activity (7 papers).
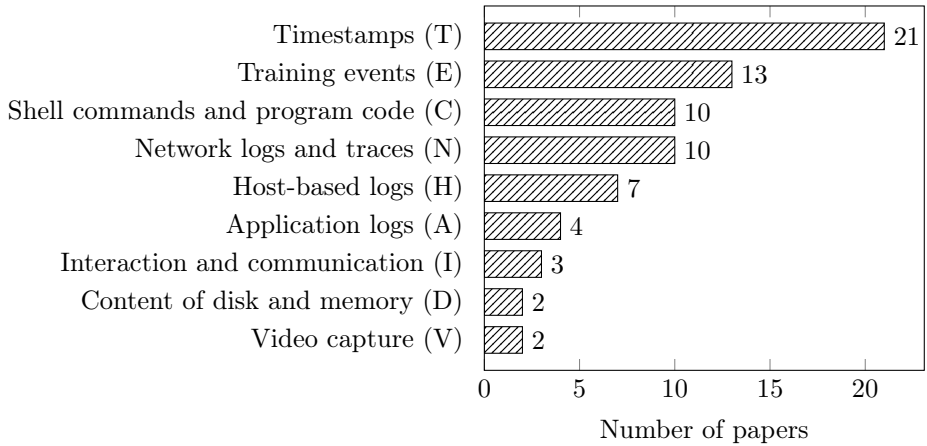
Fig. 4 The distribution of how often the identified data types were present in the 35 papers.

- *Application logs* (A), such as the status of network services, interaction with a graphical user interface, mouse clicks, and keystrokes (4 papers).
- *Content of disk and memory* (D), such as stored configuration files (2 papers).

The data generated during the training collected from LMS and other sources comprise these categories:

- *Training events* (E), that is, interactions with a LMS or actions such as submitted answers or flags in a CTF (13 papers)[1].
- *Interaction and communication* (I) between students, which includes chat content or e-mail headers (3 papers).
- *Video capture* (V) of learners' screens (2 papers).

The final category of data collected in 21 papers was *timestamps* (T) of actions such as command submissions or event triggers. This category also included the corresponding derived data, such as the duration of these actions.

In the reviewed papers, we considered only data that were collected in practice and later analyzed. For example, P15 states that their cyber range allows collecting host-based and network logs. However, the paper does not demonstrate this capability in practice. The analysis is performed only on shell commands and time-related information, so the corresponding entry for P15 in Table 2 lists only the categories C and T.

Apart from the data relevant to our research questions, some papers employed other data sources, for example, learner surveys (P8, P10, P21, P22, P24, P28, P31, P33), student grades (P12, P18, P23), observer reports (P8), sample solutions to tasks (P15, P25), and Common Weakness Enumeration (CWE) articles (P4, P29). For more additional data types, we refer the reader to the above-mentioned survey by Maennel (Maennel, 2020).

---

[1]For an in-depth overview of learning events and architecture for collecting them, see (Estévez-Ayres, Arias Fisteus, & Delgado-Kloos, 2017).

4.1.5 RQ1.5: Sample Size

The size of the collected dataset influences which EDM/LA methods are applicable. For example, many machine learning techniques require thousands or tens of thousands of data points. However, due to the diversity of data types we identified, it is impossible to compare the papers directly. For example, a dataset of 1 GB of network traffic and 1 hour of video footage of learners' screens are incomparable.

Therefore, as a proxy, we looked at the number of participants from which data were collected. The sample size ranged widely, from one (P19) to 9738 (P3, P5) participants. The median was 43. For comparison, in general cybersecurity education papers, the median number of participants is about 40 (Švábenský, Vykopal, & Čeleda, 2020).

Although most papers reported the number of participants, there were occasional issues with clarity. For example, P17 states that 900 people participated in the exercise; however, it was unclear whether all of them contributed to the dataset. In P7, 24 teams participated, but the team size is unknown[2].

4.1.6 RQ1.6: Time Span of Data

We also looked at the time span during which the data were collected since this is another proxy indicator of the dataset's depth. Most commonly, 14 papers collected the data during the period from 1 to 14 days. This was usually the case of CTF and CDX. Next, ten papers collected the data over a period from 1 to 13 hours. They mostly examined one or more lab sessions. Three papers spanned a month or more. Finally, eight papers did not report the time-related information.

4.1.7 RQ1.7: Privacy and Ethical Issues

Although almost all papers collected data about human participants, only eight publications explicitly addressed privacy and ethical issues. P3 explains that no information about individual students was recorded. Similarly, P22 and P28 explain that the collected data were not linked to personally identifiable information during the research. In P21, the data were anonymized, and P17 argues that only aggregate data are presented to preserve anonymity. Finally, P6 and P29 describe that participants explicitly consented to data collection, and in P8, the participants could opt out of the data collection.

Ethical measures, such as data anonymization, may be overlooked when reporting EDM/LA research. However, they constitute an important part of the research process, so EDM/LA researchers should not neglect them in future work.

---

[2]When computing the median of participants, we performed a small simplification for P7: by assuming two or three people per team, we estimated 60 participants in the 24 teams.

4.1.8 RQ1.8: Analysis Methods

EDM/LA offer a multitude of techniques and methods for the analysis of collected data to achieve an educational goal. We observed these six types of analysis, which are summarized in Table 2 and Figure 5.
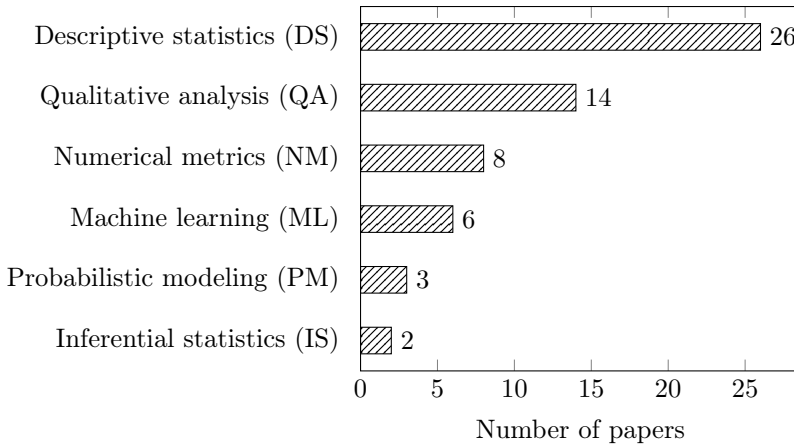

Fig. 5 The distribution of how often the identified analysis methods were present in the 35 papers.

- *Descriptive statistics* (DS) of collected data, including their correlations. Almost all papers (26 out of 35) reported some descriptive statistics.
- *Inferential statistics* (IS), which was present only in P33 and P35. Although P8, for example, applied statistical testing to questionnaire data, these data were out of the scope of this SLR. Nevertheless, since the median sample across all selected papers was 43 participants, many statistical tests might not have been appropriate in other papers.
- *Numerical metrics* (NM) computed by aggregating the collected data, usually for assessment. Eight papers employed standard metrics, such as cosine similarity, or developed custom scoring metrics.
- *Machine learning* (ML) methods, more specifically:
  - regression (P10, P21);
  - classification using Naive Bayes (P12), decision trees (P12, P24), or support vector machines (P24);
  - principal components analysis (P1); and
  - custom models (P30).
- *Probabilistic modeling* (PM) in P26, P29, and P35.
- *Qualitative analysis* (QA) of the collected data, either of the submitted commands (P7, P13, P19) or other student actions (P11, P17, P20, P28, P34). Moreover, six other papers (P5, P21, P22, P25, P29, P35) used qualitative analysis in addition to other methods.

Most data analyses were performed after the training ended, which seems to be the most straightforward use case. The exceptions included live scoring during a CDX (for example, P8 and P16).

4.2 Research Topic 2: Research Impact

Next, we reviewed the papers' application domain, contribution, supplementary materials, and citation impact.

4.2.1 RQ2.1: Application in Practice

The published research had diverse application contexts, which again demonstrates that EDM/LA is suitable for various use cases. 13 papers were applied within university courses. Next, five papers were from a CDX and five more from a jeopardy CTF. Ten were applied in other types of cybersecurity training. Finally, two papers had no application besides the author testing.

4.2.2 RQ2.2: Contribution

Table 2 shows an overview of the contribution of the individual papers. In this section, we discuss the specific contribution of each paper in more detail. The papers are grouped by the application context.

The majority of papers used student interaction data within a university course. P7 and P13 generated graphical progress models of students' approaches to solving cybersecurity exercises using a command line. P12 and P14 used mainly network logs to assess students in cybersecurity courses. P18 analyzed whether introducing an optional story element into cybersecurity assignments improves student performance. P23 and P24 assessed student learning mainly based on timing information, such as time spent on the tasks. Similarly, P26, P28, P31, and P33 measured skills of students using time-base statistics and event logs, in addition. P27 evaluated students' network configuration and P32 students' usage of unsafe programming functions.

A compact cluster of papers concentrated on CDX. P8 collected network and system logs to study the performance of participating teams. Similar data sources were used in P10 to assess and predict team performance. P17 proposed a more systematic approach: a methodology to employ CDX data for team assessment. P16 focused on using CDX data to provide feedback for Red teams. Finally, P20 developed a tool to analyze Blue team communication and reporting data.

Jeopardy CTF was another important application area. P2 collected learner interaction data from a CTF platform to measure score distribution, time delays, and frustration of participants. P6 followed up on this work to derive meaningful blocks of learner activity, such as the used applications, from CTF logs. P3 analyzed challenge completion in a large-scale online Jeopardy CTF. P5 used the same dataset as P3 to report on observed cheating attempts and proposed a method called automatic problem generation as a solution. Finally, P11 used host-based logs to describe learner activity.

Most of the remaining works focused on other types of cybersecurity training. P1 proposed an evidence model for analyzing data from Packet Tracer, software for learning networking, to discover skill levels of learners. P9, P21, and P22 evaluated data capturing learner interactions with a cyber range to understand how students approach solving cybersecurity exercises. P15 employed metrics such as timing, commands entered, and similarity to the reference solution to assess technical skills of trainees in a cyber range. P25 generated visual models of trainees' approaches and investigated their difference from the reference solution, again using this information for skill assessment. P29 and P30 created sophisticated numerical and machine learning models to analyze log data from cybersecurity training. P34 used operation logs to evaluate a cybersecurity course, and P35 focused on student collaboration in group programming tasks.

Finally, two papers presented only a prototype. P4 proposed a method for analyzing program code to discover vulnerabilities and recommend relevant sources to software developers. Lastly, P19 proposed a method for real-time analysis of log data from cyber ranges to improve situational awareness.

In a few cases, however, the contribution was difficult to determine. For example, the abstract of P11 states that the data analysis will be used to reveal student misconceptions. However, no reported results indicated the misconceptions were found.

4.2.3 RQ2.3: Supplementary Materials

Only eight publications provided supplementary materials along with the paper. Papers about the PicoCTF platform (P3, P5) released the open-source code of the platform on GitHub. The Frankenstack framework (P16) can also be found on GitHub, although the repository is not linked from the paper. Similarly, P21 links a repository with an open-source visualization tool, and P29 refers to open-source components of the training platform. P18 provides virtual machines and cybersecurity exercises for other instructors. P28 also links exercises, but the link is no longer functional. Finally, P33 provides a video about the training platform and solving exercise tasks in it.

4.2.4 RQ2.4: Citation Analysis

The citation analysis was conducted on October 8, 2021 using Scopus. Although the citation counts are relatively small (min = 0, max = 64, median = 5), this is probably because 71% of the papers were published in 2017 or later. As a result, there was not enough time for the citation impact to appear. However, the sample is too small for conclusive results.

Interestingly, although the most cited journal paper (P19) has 64 citations, the number drops to 32 after removing self-citations. On the other hand, the most cited conference paper (P3) has 63 total citations and 62 non-self-citations. This paper deals with PicoCTF, a popular event held annually since 2013, along with an associated open-source platform.

4.3 Research Topic 3: Research Community

We now present the results regarding the authors, their affiliations, and publication venues.

4.3.1 RQ3.1: Authors and Their Affiliations

A total of 125 unique authors wrote the surveyed papers. Out of these authors, 101 co-authored only one paper, 21 co-authored two papers, and 2 authors three papers. This suggests that the community of EDM/LA researchers in cybersecurity is neither stable nor particularly big.

Considering the authors' affiliations, 92 of them were associated with a university or a college, 18 with a military or government institution, 13 with a private company, and 4 with a non-governmental research and development organization[3]. The prevalence of academic institutions is motivated by the fact that the authors often work as teachers of cybersecurity courses, and the research supports their teaching. Another contributing factor is that their institutions may require them to publish as a part of their job duties.

Of the 35 selected papers, 22 were written solely by university researchers and 3 by military/government institutions. We observed little cross-institutional collaboration. Universities and military collaborated in 3 cases, universities and private companies in 3 cases, and research institutes and private companies in 4 papers. Such collaboration can be beneficial, because the research addresses the needs of various stakeholders, and the educational intervention is evaluated at multiple institutions.

4.3.2 RQ3.2: Publication Venues

The selected papers were published in various conferences and journals; there were no prominent flagship venues for EDM/LA research in cybersecurity. However, some trends appeared: conferences are preferred to journals (27 vs. 8), probably due to the speed of publication and targeting a specific audience. Also, bibliometrics is not an important criterion for most authors. Half of the journal papers were not indexed in Web of Science (Clarivate, 2022), and most conferences were not CORE-ranked (Computing Research and Education Association of Australasia, 2021). However, other metrics and standards for rating the quality of publication venues also exist, so these provide only a partial point of view.

4.3.3 RQ3.3: Citation Map

We examined whether there are citation interconnections between papers that might indicate relationships between researchers. However, as Figure 6 shows, the papers rarely cite each other. The most cited paper is P10 with three non-self citations from P11, P16, and P17, indicating that P10 may represent important prior work. Next, P3 has two citations by P18 and P22, both of

---

[3]The counts sum to 127 because two authors had two affiliations.

which deal with offensive cybersecurity topics similar to P3. P5 is cited by P31, again overlapping in topics. Lastly, P2 and P8 have a single citation each, both from P17, focusing on defensive topics. Overall, these links are relatively weak, showing that the community is fragmented, with the only prominent group forming in the CDX application domain.
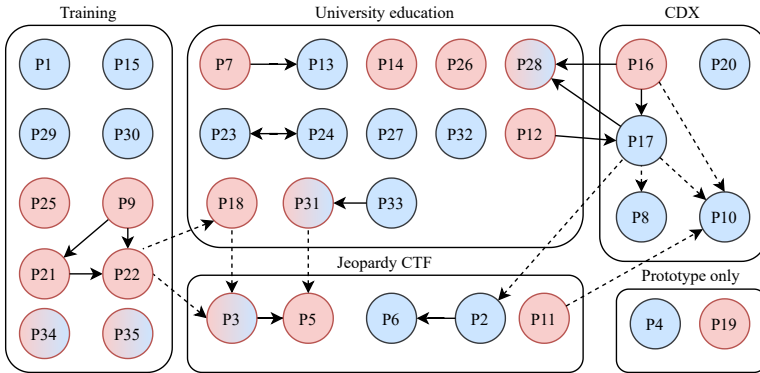

Fig. 6 The 35 papers grouped by the application context (RQ2.1). Red nodes deal with offensive topics and blue nodes with defensive (RQ1.1). A solid arrow indicates overlapping authors and time progression ($A \rightarrow B$ means that B was published after A and shared one or more authors). A dashed arrow indicates non-self citations ($B \dashrightarrow A$ means that B cites a previously published paper A and both have different authors).

4.4 Summary of the Observed Trends

We now summarize the key results of the literature review. The papers cover offensive and defensive topics almost equally, which can also be seen in Figure 6. The authors are usually from universities or military institutions, and their typical goal is to assess students. To do so, the researchers collect data from the training infrastructure, which is often virtualized.

When sorted by the most common data type, the data include timestamped actions in a learning infrastructure, training events, commands and code, network traces, and host-based logs. They are collected from a median of 43 students over a period of a few hours to several days. To analyze the data, the authors usually present descriptive statistics or qualitative insights.

The research often provides valuable contributions, such as innovative scoring methods, insights into students' behavior, or case studies carried out in an authentic context. As Figure 6 shows, the application context is relatively evenly distributed among university education, CDX, CTF, and other types of cybersecurity training.

The papers' main disadvantage is that they rarely provide supplementary materials, preventing other researchers and educators from building on the results. This is also true in other cybersecurity education papers (Švábenský,

Vykopal, & Čeleda, 2020). Moreover, few of them address the privacy of the students and data anonymization issues.

4.5 Limitations of the Results

All literature reviews are limited by the selection of paper databases and search terms. Although we focused only on papers indexed by Scopus, it is a major database, and our search query was broad and reviewed by independent cybersecurity experts. Therefore, we believe we minimized the number of missed candidate papers.

Another limitation of all surveys is a potential researcher bias when manually reviewing the papers and extracting data from them (Petersen et al., 2015). Nevertheless, we minimized this bias by following the guidelines for literature reviews (Kitchenham & Charters, 2007; Petersen et al., 2008, 2015; Randolph, 2009). Most notably, these include: defining the SLR protocol in advance, having two authors review the candidate papers independently, and discussing and resolving disagreements.

Even though our search criteria were broad, we selected only 35 papers. This number suggests that the analysis of cybersecurity training artifacts is a narrow domain, and its community is not (yet) widely established. To support the development of this arising research area, we provide recommendations for further research in Section 5. Nevertheless, the relatively small number of selected papers is typical for SLRs, which tend to have a narrow focus. In a comparison of ten literature reviews (Petersen et al., 2008, p. 7), five of them inspected less than 30 papers.

5 Implications of this Literature Survey

Cybersecurity education research combined with EDM/LA has a substantial practical impact. It brings new insights into learning technologies and enables effective learning interventions. However, our SLR revealed that few studies fully exploit the potential of EDM/LA applied to student-generated data. One of the reasons could be that creating the training content itself is challenging, and few resources are left for other activities such as follow-up research.

To motivate further research, we formulate a research agenda in Section 5.1. Section 5.2 also provides recommendations for writing papers in the domain of EDM/LA to aid fellow cybersecurity education researchers. As a result, this SLR not only reviews the facts derived from existing literature but adds to the understanding of the area.

5.1 Identified Research Gaps and Future Work Proposals

We list several open problems not covered in the surveyed papers. To be specific, we phrase the problems as research questions and invite interested researchers to address them.

1. The vast majority of the examined papers focused on offensive security and network security. An uncovered research question is: *How can student data be leveraged to support other areas of cybersecurity education, such as secure programming, data security, or even human security?*

2. Most reviewed papers performed a post-hoc analysis of student data. However, a real-time analysis would provide situational awareness and support classroom orchestration. Moreover, it would enable providing immediate automated feedback to students to improve their learning experience. Therefore, an interesting question is: *Which information can be inferred from student data during the training to inform instructors and students about their progress?*

3. A follow-up question to the previous one is: *How to automatically adapt instruction based on the student data?* Similarly to intelligent tutoring systems, cybersecurity training environments can employ student data to personalize instruction according to the skill level of individual students. These technologies can reduce the barrier to participation of beginners.

4. Since the median sample size was only 43 students, many statistical and machine learning methods are not applicable. However, the time span of data collection ranged from several hours to several days, during which each student generates in-depth data. So, it would be interesting to examine: *Which automated methods for data analysis are suitable for a small number of students who interact with the training environment for a long time?*

5. For researchers interested in writing literature survey papers, a relevant question is: *How can cybersecurity be taught?* The review can examine the possible teaching methods, their effectiveness, advantages and disadvantages, and necessary infrastructure.

Moreover, open problems in computing education research (Denny, Becker, Craig, Wilson, & Banaszkiewicz, 2019) or general cybersecurity education (Švábenský, Vykopal, & Čeleda, 2020) can also be applied to hands-on security training combined with EDM/LA.

Last but not least, cybersecurity education research requires an infrastructure for data collection. To support the research, developers of learning technologies can examine how to simplify the deployment of training environments that would enable seamless data collection and analysis.

Overall, there are many opportunities for fruitful future work. Research aimed at technologies that support learning has a great potential to improve the student experience. It can enable remote access to education at scale, provide rapid assessment and feedback, and reduce the burden placed on instructors.

5.2 Recommendations for Publishing EDM/LA Research

Handbooks of EDM (Romero et al., 2010) and LA (Lang et al., 2017) provide an excellent overview of general methods and research approaches, along with examples of studies. We wish to add more specific recommendations that we

formulated while reviewing the 35 papers. Therefore, we provide a list of six criteria that an EDM/LA paper in cybersecurity should address.

1. Clearly describe the learning objectives of the cybersecurity educational intervention. Refer to a standardized curriculum, such as CSEC2017 (Joint Task Force on Cybersecurity Education, 2017) or ACM/IEEE curricular guidelines (CC2020 Task Force, 2020; Joint Task Force on Computing Curricula, Association for Computing Machinery (ACM) and IEEE Computer Society, 2013).

2. Follow a thorough methodology for data collection.
   - Explain the purpose of the data collection. Usually, this is bound to the studied research question. Are the data needed to assess students, help them learn, provide feedback to the training designers, or something else?
   - Characterize the technical environment from which the data are collected. Is it a physical or virtual infrastructure? What would other researchers or instructors need to build it? Does it employ open-source components?
   - List what data are collected, from which systems, and how they address the defined purpose.
   - State the precise number of students (participants) from which the data were collected.
   - Report the time span during which the data were collected.
   - Explain the ethical measures. Which steps were taken to anonymize the data and preserve the privacy of participants? This usually involves informed consent or approval of the study by an institutional review board.

3. Select appropriate analysis methods suitable for the collected data and relevant for the defined purpose.

4. Explicitly describe the contributions and practical applications of the research.

5. When applying the research to practice, evaluate that the EDM/LA interventions helped improve some aspect of teaching or learning.

6. If possible, publish the dataset, source code, configuration files, other relevant supplementary materials, and documentation on using or deploying them. This enables other researchers and instructors to replicate the setup and build upon the results.

To increase the quality of cybersecurity education research papers, we also recommend the authors to study the methodology of computing education research (Fincher & Robins, 2019; Lishinski, Good, Sands, & Yadav, 2016). For the authors who want to apply inferential statistics to their data, we recommend the review of usage of statistics in computing education research (Sanders et al., 2019). Finally, all authors who present analysis results visually can benefit from suggestions in (Simon et al., 2019).

Regarding the publication of datasets, open-data initiatives such as Zenodo (CERN Data Centre & Invenio, 2022) allow researchers to easily and

permanently share their research data. When looking at specific communities of practice, activities such as The Graphics Replicability Stamp Initiative (The Graphics Replicability Stamp Initiative, 2017) encourage researchers in the field of computer graphics and visualizations to publish their research artifacts. A similar initiative would benefit the cybersecurity education community as well. It is challenging to deploy infrastructure for the training itself, not to mention enhancing it with data collection capabilities. Researchers who overcome this issue can publish their dataset and enable others to analyze it.

6 Conclusions

We surveyed publications that leverage student-generated data to support hands-on cybersecurity training. This interdisciplinary research area is still in its developing stages, so it is not yet well-understood. Our work helps to contextualize it and provides inspiration for researchers, practitioners, and educators.

We followed the best practices for systematic literature reviews, revealing diverse applications of EDM/LA methods on training artifacts. These range from understanding students' misconceptions, to developing tools for providing automated feedback, to evaluating assessment models for skill level prediction. As a result, EDM/LA research yields insights beneficial for students, instructors, and developers of interactive learning environments.

The emerging research in this area will also have a practical impact. Cybersecurity is a domain that needs educated experts – millions of skilled workers now lack worldwide ($(ISC)^2$, 2021). Examining new ways of employing student interaction data will enable a better understanding of teaching and learning processes and, ultimately, improve them. As a result, instructors will be better equipped to train cybersecurity specialists more efficiently. What is more, the surveyed methods are applicable in other domains of computing education, such as operating systems, networking, or programming.

Although our survey showed that student data from hands-on training have vast potential, researchers do not (yet) fully exploit it. We hope to support future research efforts in this area by providing the following contributions:

- An organized inventory of papers along with the synthesis and classification of their approaches, results, and applications. This inventory evaluates the current trends and adds to the understanding of the area. It will aid both new researchers and those already in the field and inspire developers of cybersecurity learning technologies, instructors, and creators of training content.
- Identification of research trends as well as gaps, along with potential directions for future work to motivate further research.
- Practical recommendations for conducting cybersecurity education research.

As supplementary material (Švábenský, Vykopal, Čeleda, & Kraus, 2022) for the paper, we publish the raw dataset: BibTeX references exported from

Mendeley reference manager (Mendeley, 2021) that include citations of all candidate and selected papers. We also provide the processed dataset: a spreadsheet with the complete information about the 35 selected papers.

Statements and Declarations

Competing interests.   The authors have no competing interests to declare.

Availability of data and materials.   The accompanying data and materials are published in a free, open-source repository on Zenodo (Švábenský et al., 2022).

Authors' contributions.   *Valdemar Švábenský*: Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Visualization, Project administration. *Jan Vykopal*: Conceptualization, Methodology, Investigation, Writing – Review & Editing, Data Curation. *Pavel Čeleda*: Conceptualization, Writing – Review & Editing, Supervision, Funding acquisition. *Lydia Kraus*: Investigation, Writing – Review & Editing

References

Abbott, R.G., McClain, J., Anderson, B., Nauer, K., Silva, A., Forsythe, C. (2015). Log Analysis of Cyber Security Training Exercises. *Procedia Manufacturing*, *3*, 5088–5094. Retrieved from https://doi.org/10.1016/j.promfg.2015.07.523

Almansoori, M., Lam, J., Fang, E., Mulligan, K., Soosai Raj, A.G., Chatterjee, R. (2020). How Secure Are Our Computer Systems Courses? *Proceedings of the 2020 ACM Conference on International Computing Education Research* (p. 271–281). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3372782.3406266

Andreatos, A.S. (2017). Designing educational scenarios to teach network security. *8th IEEE Global Engineering Education Conference, EDUCON 2017* (pp. 1606–1610). Washington, D.C., USA: IEEE Computer Society. Retrieved from https://doi.org/10.1109/EDUCON.2017.7943063

Andreolini, M., Colacino, V.G., Colajanni, M., Marchetti, M. (2019). A Framework for the Evaluation of Trainee Performance in Cyber Range Exercises. *Mobile Networks and Applications*, *25*, 236–247. Retrieved from https://doi.org/10.1007/s11036-019-01442-0

Burket, J., Chapman, P., Becker, T., Ganas, C., Brumley, D. (2015). Automatic Problem Generation for Capture-the-Flag Competitions. *2015 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 15)* (pp. 1–8). Berkeley, CA, USA: USENIX Association. Retrieved from https://www.usenix.org/conference/3gse15/summit-program/presentation/burket

Caliskan, E., Tatar, U., Bahsi, H., Ottis, R., Vaarandi, R. (2017). Capability detection and evaluation metrics for cyber security lab exercises. *Proceedings of the International Conference on Cyber Warfare and Security* (pp. 407–414). Sonning Common, Reading, UK: Academic Conferences and Publishing International.

CC2020 Task Force (2020). *Computing Curricula 2020: Paradigms for Global Computing Education.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3467967

CERN Data Centre & Invenio (2022). *Zenodo – Research. Shared.* Retrieved April 6, 2022 from https://zenodo.org.

Chapman, P., Burket, J., Brumley, D. (2014). PicoCTF: A Game-Based Computer Security Competition for High School Students. *2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14)* (pp. 1–10). Berkeley, CA, USA: USENIX Association. Retrieved from https://www.usenix.org/conference/3gse14/summit-program/presentation/chapman

Chothia, T., Holdcroft, S., Radu, A.-I., Thomas, R.J. (2017). Jail, Hero or Drug Lord? Turning a Cyber Security Course Into an 11 Week Choose Your Own Adventure Story. *2017 USENIX Workshop on Advances in Security Education (ASE 17)* (pp. 1–11). Berkeley, CA, USA: USENIX Association. Retrieved from https://www.usenix.org/conference/ase17/workshop-program/presentation/chothia

Chothia, T., & Novakovic, C. (2015). An Offline Capture The Flag-Style Virtual Machine and an Assessment of Its Value for Cybersecurity Education. *2015 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 15)* (pp. 1–8). Berkeley, CA, USA: USENIX Association. Retrieved from https://www.usenix.org/conference/3gse15/summit-program/presentation/chothia

Clarivate (2022). *InCites Journal Citation Reports.* Retrieved April 6, 2022 from https://jcr.clarivate.com/jcr/home.

Computing Research and Education Association of Australasia (2021). *CORE.* Retrieved April 6, 2022 from http://portal.core.edu.au/conf-ranks/.

Cybersecurity & Infrastructure Security Agency (2018). *Cyber Storm: Securing Cyber Space.* Retrieved April 6, 2022 from https://www.cisa.gov/cyber-storm-securing-cyber-space.

DEF CON (2021). *CTF Archive.* Retrieved April 6, 2022 from https://www.defcon.org/html/links/dc-ctf.html.

Deng, Y., Lu, D., Chung, C.-J., Huang, D., Zeng, Z. (2018). Personalized Learning in a Virtual Hands-on Lab Platform for Computer Science Education. *2018 IEEE Frontiers in Education Conference (FIE)* (pp. 1–8). New York, NY, USA: IEEE. Retrieved from https://doi.org/10.1109/FIE.2018.8659291

Denny, P., Becker, B.A., Craig, M., Wilson, G., Banaszkiewicz, P. (2019). Research This! Questions That Computing Educators Most Want Computing Education Researchers to Answer. *Proceedings of the 2019 ACM Conference on International Computing Education Research* (p. 259–267). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3291279.3339402

Elsevier (2021). *Scopus.* Retrieved April 6, 2022 from https://www.scopus.com.

Espinha Gasiba, T., Lechner, U., Pinto-Albuquerque, M. (2020, 11). Cybersecurity Challenges in Industry: Measuring the Challenge Solve Time to Inform Future Challenges. *Information*, *11*(11), 533. Retrieved from https://doi.org/10.3390/info11110533

Estévez-Ayres, I., Arias Fisteus, J., Delgado-Kloos, C. (2017). Lostrego: A distributed stream-based infrastructure for the real-time gathering and analysis of heterogeneous educational data. *Journal of Network and Computer Applications*, *100*, 56-68. Retrieved from https://doi.org/10.1016/j.jnca.2017.10.014

Falah, A., Pan, L., Chen, F. (2019). A Quantitative Approach to Design Special Purpose Systems to Measure Hacking Skills. *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)* (pp. 54–61). New York, NY, USA: IEEE. Retrieved from https://doi.org/10.1109/TALE.2018.8615431

Fincher, S.A., & Robins, A.V. (Eds.). (2019). *The Cambridge Handbook of Computing Education Research.* Cambridge, United Kingdom: Cambridge University Press. Retrieved from https://doi.org/10.1017/9781108654555

Google (2021). *Capture the Flag.* Retrieved April 6, 2022 from https://capturetheflag.withgoogle.com.

Graham, K., Anderson, J., Rife, C., Heitmeyer, B., R. Patel, P., Nykl, S., ... D. Merkle, L. (2020). Cyberspace Odyssey: A Competitive Team-Oriented Serious Game in Computer Networking. *IEEE Transactions on Learning Technologies*, *13*(3), 502-515. Retrieved from https://doi.org/10.1109/TLT.2020.3008607

Granåsen, M., & Andersson, D. (2016). Measuring team effectiveness in cyber-defense exercises: a cross-disciplinary case study. *Cognition, Technology & Work*, *18*(1), 121–143. Retrieved from https://doi.org/10.1007/s10111-015-0350-2

Henshel, D.S., Deckard, G.M., Lufkin, B., Buchler, N., Hoffman, B., Rajivan, P., Collman, S. (2016). Predicting proficiency in cyber defense team exercises. *MILCOM 2016 – IEEE Military Communications Conference* (pp. 776–781). New York, NY, USA: IEEE. Retrieved from https://doi.org/10.1109/MILCOM.2016.7795423

Hundhausen, C., Olivares, D., Carter, A. (2017, 08). IDE-Based Learning Analytics for Computing Education: A Process Model, Critical Review, and Research Agenda. *ACM Transactions on Computing Education*, *17*(3), 11:1–11:26. Retrieved from https://doi.org/10.1145/3105759

Ihantola, P., Vihavainen, A., Ahadi, A., Butler, M., Börstler, J., Edwards, S.H., ... Toll, D. (2015). Educational Data Mining and Learning Analytics in Programming: Literature Review and Case Studies. *Proceedings of the 2015 ITiCSE on Working Group Reports* (pp. 41–63). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2858796.2858798

Imani, M., & Montazer, G.A. (2019). A survey of emotion recognition methods with emphasis on e-learning environments. *Journal of Network and Computer Applications*, *147*, 102423. Retrieved from https://doi.org/10.1016/j.jnca.2019.102423

(ISC)^2 (2021). *Cybersecurity Workforce Study* (Tech. Rep.). Retrieved from https://www.isc2.org/Research/Workforce-Study

Joint Task Force on Computing Curricula, Association for Computing Machinery (ACM) and IEEE Computer Society (2013). *Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science.* New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2534860

Joint Task Force on Cybersecurity Education (2017). *Cybersecurity Curricular Guideline.* Retrieved April 6, 2022 from https://cybered.acm.org.

Kaneko, K., Igarashi, T., Kayama, K., Takeuchi, T., Suzuki, T., Kawase, A., ... Okamura, K. (2020). Learning Analytics with Multi-faced Data for Cybersecurity Education. *9th International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 244–249). Retrieved from https://doi.org/10.1109/IIAI-AAI50415.2020.00055

Kasurinen, J., & Knutas, A. (2018). Publication trends in gamification: A systematic mapping study. *Elsevier Computer Science Review*, *27*, 33–44. Retrieved from https://doi.org/10.1016/j.cosrev.2017.10.003

Khando, K., Gao, S., Islam, S.M., Salman, A. (2021). Enhancing employees information security awareness in private and public organisations: A systematic literature review. *Computers & Security*, *106*, 102267. Retrieved from https://doi.org/10.1016/j.cose.2021.102267

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (Tech. Rep.). EBSE.

Knobbout, J., & Van Der Stappen, E. (2020). Where is the Learning in Learning Analytics? A Systematic Literature Review on the Operationalization of Learning-Related Constructs in the Evaluation of Learning Analytics Interventions. *IEEE Transactions on Learning Technologies*, *13*(3), 631–645. Retrieved from https://doi.org/10.1109/TLT.2020.2999970

Kokkonen, T., & Puuska, S. (2018). Blue Team Communication and Reporting for Enhancing Situational Awareness from White Team Perspective in Cyber Security Exercises. *18th International Conference on Next Generation Teletraffic and Wired/Wireless Advanced Networks and Systems, NEW2AN 2018 and 11th Conference on Internet of Things and Smart Spaces, ruSMART 2018* (Vol. 11118 LNCS, pp. 277–288). Vienna, Austria: Springer. Retrieved from https://doi.org/10.1007/978-3-030-01168-0_26

Kont, M., Pihelgas, M., Maennel, K., Blumbergs, B., Lepik, T. (2017). Frankenstack: Toward real-time Red Team feedback. *2017 IEEE Military*

Communications Conference, MILCOM 2017 (pp. 400–405). New York, NY, USA: IEEE. Retrieved from https://doi.org/10.1109/MILCOM .2017.8170852

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research*, *30*(3), 411–433. Retrieved from https://doi.org/10.1111/j.1468-2958 .2004.tb00738.x

Kucek, S., & Leitner, M. (2020). An Empirical Survey of Functions and Configurations of Open-Source Capture the Flag (CTF) Environments. *Journal of Network and Computer Applications*, *151*. Retrieved from https://doi.org/10.1016/j.jnca.2019.102470

Labuschagne, W.A., & Grobler, M. (2017). Developing a capability to classify technical skill levels within a cyber range. *16th European Conference on Cyber Warfare and Security, ECCWS 2017* (pp. 224–234). Red Hook, NY, USA: Curran Associates Inc. Retrieved from https://www.proquest .com/docview/1966803837

Lang, C., Siemens, G., Wise, A., & Gašević, D. (Eds.). (2017). *Handbook of learning analytics* (1st ed.). Society for Learning Analytics Research (SoLAR). Retrieved from https://doi.org/10.18608/hla17

Liñán, L.C., & Pérez, Á.A.J. (2015). Educational data mining and learning analytics: differences, similarities, and time evolution. *International Journal of Educational Technology in Higher Education*, *12*(3), 98–112. Retrieved from https://doi.org/10.7238/rusc.v12i3.2515

Lishinski, A., Good, J., Sands, P., Yadav, A. (2016). Methodological Rigor and Theoretical Foundations of CS Education Research. *Proceedings of the 2016 ACM Conference on International Computing Education Research* (p. 161–169). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2960310.2960328

Luxton-Reilly, A., Simon, Albluwi, I., Becker, B.A., Giannakos, M., Kumar, A.N., . . . Szabo, C. (2018). Introductory Programming: A Systematic Literature Review. *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (pp. 55–106). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3293881.3295779

Maennel, K. (2020). Learning Analytics Perspective: Evidencing Learning from Digital Datasets in Cybersecurity Exercises. *2020 IEEE European*

Symposium on Security and Privacy Workshops (EuroSPW) (pp. 27–36). Retrieved from https://doi.org/10.1109/EuroSPW51379.2020.00013

Maennel, K., Mäses, S., Sütterlin, S., Ernits, M., Maennel, O. (2019). Using technical cybersecurity exercises in university admissions and skill evaluation. *IFAC-PapersOnLine*, *52*(19), 169-174. Retrieved from https://doi.org/10.1016/j.ifacol.2019.12.169   (14th IFAC Symposium on Analysis, Design, and Evaluation of Human Machine Systems (HMS 2019))

Maennel, K., Ottis, R., Maennel, O. (2017). Improving and measuring learning effectiveness at cyber defense exercises. *22nd Nordic Conference on Secure IT Systems, NordSec 2017* (pp. 123–138). Vienna, Austria: Springer. Retrieved from https://doi.org/10.1007/978-3-319-70290-2_8

Malmi, L., Sheard, J., Simon, Bednarik, R., Helminen, J., Korhonen, A., . . . Taherkhani, A. (2010). Characterizing Research in Computing Education: A Preliminary Analysis of the Literature. *Proceedings of the Sixth International Workshop on Computing Education Research* (pp. 3–12). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1839594.1839597

Mangaroska, K., & Giannakos, M. (2019). Learning Analytics for Learning Design: A Systematic Literature Review of Analytics-Driven Design to Enhance Learning. *IEEE Transactions on Learning Technologies*, *12*(4), 516–534. Retrieved from https://doi.org/10.1109/TLT.2018.2868673

Margulieux, L., Ketenci, T.A., Decker, A. (2019). Review of measurements used in computing education research and suggestions for increasing standardization. *Computer Science Education*, *29*(1), 49–78. Retrieved from https://doi.org/10.1080/08993408.2018.1562145

Matcha, W., Uzir, N.A., Gašević, D., Pardo, A. (2020). A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. *IEEE Transactions on Learning Technologies*, *13*(2), 226–245. Retrieved from https://doi.org/10.1109/TLT.2019.2916802

Mendeley (2021). *Reference Manager.* Retrieved April 6, 2022 from https://www.mendeley.com/reference-management/reference-manager.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Group, T.P. (2009, 07). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*, *6*(7), 1-6. Retrieved from

https://doi.org/10.1371/journal.pmed.1000097

Nadeem, M., Allen, E.B., Williams, B.J. (2015). A Method for Recommending Computer-Security Training for Software Developers: Leveraging the Power of Static Analysis Techniques and Vulnerability Repositories. *12th International Conference on Information Technology – New Generations* (pp. 534–539). New York, NY, USA: IEEE. Retrieved from https://doi.org/10.1109/ITNG.2015.90

Natural Language Toolkit (NLTK) Project (2022). *Source code for* `nltk.metrics.agreement`. Retrieved April 6, 2022 from http://www.nltk.org/_modules/nltk/metrics/agreement.html.

Nunn, S.G., Avella, J.T., Kanai, T., Kebritchi, M. (2016). Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review. *Online Learning*, *20*(2), 13–29. Retrieved from https://doi.org/10.24059/olj.v20i2.790

Palmer, N. (2019). Automating the Assessment of Network Security in Higher Education. *2019 International Conference on Computing, Electronics Communications Engineering (iCCECE)* (pp. 141–146). Retrieved from https://doi.org/10.1109/iCCECE46942.2019.8941804

Papamitsiou, Z., Giannakos, M., Simon, Luxton-Reilly, A. (2020). Computing Education Research Landscape through an Analysis of Keywords. *Proceedings of the 2020 ACM Conference on International Computing Education Research* (pp. 102–112). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3372782.3406276

Petersen, K., Feldt, R., Mujtaba, S., Mattsson, M. (2008). Systematic Mapping Studies in Software Engineering. *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering* (pp. 68–77). Swindon, UK: BCS Learning & Development Ltd. Retrieved from https://dl.acm.org/doi/10.5555/2227115.2227123

Petersen, K., Vakkalanka, S., Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, *64*, 1–18. Retrieved from https://doi.org/10.1016/j.infsof.2015.03.007

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, *41*(4, Part 1), 1432–1462. Retrieved from https://doi.org/10.1016/j.eswa.2013

.08.042

Randolph, J. (2009). A guide to writing the dissertation literature review. *Practical Assessment, Research, and Evaluation*, *14*(1), 13. Retrieved from https://doi.org/10.7275/b0az-8t74

Reed, T., Nauer, K., Silva, A. (2013). Instrumenting competition-based exercises to evaluate cyber defender situation awareness. *International Conference on Augmented Cognition* (pp. 80–89). Vienna, Austria: Springer. Retrieved from https://doi.org/10.1007/978-3-642-39454-6_9

Rege, A., Obradovic, Z., Asadi, N., Parker, E., Masceri, N., Singer, B., Pandit, R. (2017). Using a Real-Time Cybersecurity Exercise Case Study to Understand Temporal Characteristics of Cyberattacks. *Social, cultural, and behavioral modeling* (pp. 127–132). Cham, Switzerland: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-60240-0_16

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R.S. (Eds.). (2010). *Handbook of educational data mining*. Boca Raton, FL, USA: CRC Press. Retrieved from https://doi.org/10.1201/b10274

Rupp, A.A., Levy, R., Dicerbo, K.E., Sweet, S.J., Crawford, A.V., Calico, T., ... Behrens, J.T. (2012). Putting ECD into practice: The interplay of theory and data in evidence models within a digital learning environment. *Journal of Educational Data Mining (JEDM)*, *4*(1), 49–110. Retrieved from https://doi.org/10.5281/zenodo.3554643

Sanders, K., Sheard, J., Becker, B.A., Eckerdal, A., Hamouda, S., Simon. (2019). Inferential Statistics in Computing Education Research: A Methodological Review. *Proceedings of the 2019 ACM Conference on International Computing Education Research* (p. 177–185). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3291279.3339408

Sheng, Q.-w. (2020). Effectiveness Evaluation of Network Security Knowledge Training Based on Machine Learning. S. Liu, G. Sun, & W. Fu (Eds.), *e-Learning, e-Education, and Online Training* (pp. 25–37). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-030-63955-6_3

Simon, S., Becker, B.A., Hamouda, S., McCartney, R., Sanders, K., Sheard, J. (2019). Visual Portrayals of Data and Results at ITiCSE. *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer*

Science Education (p. 51–57). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3304221 .3319742

Taylor, C., Arias, P., Klopchic, J., Matarazzo, C., Dube, E. (2017). CTF: State-of-the-art and building the next generation. *2017 USENIX Workshop on Advances in Security Education (ASE 17).* USENIX Association. Retrieved from https://www.usenix.org/conference/ase17/ workshop-program/presentation/taylor

The Graphics Replicability Stamp Initiative (2017). *GRSI.* Retrieved April 6, 2022 from http://www.replicabilitystamp.org.

The NATO Cooperative Cyber Defence Centre of Excellence (2021a). *Crossed Swords.* Retrieved April 6, 2022 from https://ccdcoe.org/exercises/ crossed-swords.

The NATO Cooperative Cyber Defence Centre of Excellence (2021b). *Locked Shields.* Retrieved April 6, 2022 from https://ccdcoe.org/exercises/ locked-shields.

Tian, Z., Cui, Y., An, L., Su, S., Yin, X., Yin, L., Cui, X. (2018). A real-time correlation of host-level events in cyber range service for smart campus. *IEEE Access*, *6*, 35355–35364. Retrieved from https://doi.org/10.1109/ ACCESS.2018.2846590

Tobarra, L., Robles-Gómez, A., Pastor, R., Hernández, R., Duque, A., Cano, J. (2020, 02). Students' acceptance and tracking of a new container-based virtual laboratory. *Applied Sciences*, *10*(3). Retrieved from https:// doi.org/10.3390/app10031091

Tobarra, L., Trapero, A.P., Pastor, R., Robles-Gómez, A., Hernández, R., Duque, A., Cano, J. (2020). Game-based learning approach to cybersecurity. *2020 IEEE Global Engineering Education Conference (EDUCON)* (p. 1125-1132). Retrieved from https://doi.org/10.1109/EDUCON45650 .2020.9125202

Tseng, S.-S., Lin, S.-C., Mao, C.-H., Lee, T.-J., Qiu, G.-W., Lin, M.-H. (2017, 08). An ontology guiding assessment framework for hacking competition. *2017 10th International Conference on Ubi-media Computing and Workshops (Ubi-Media)* (pp. 1–4). New York, NY, USA: IEEE. Retrieved from https://doi.org/10.1109/UMEDIA.2017.8074131

Vigna, G., Borgolte, K., Corbetta, J., Doupé, A., Fratantonio, Y., Invernizzi, L., . . . Shoshitaishvili, Y. (2014). Ten Years of iCTF: The Good, The

Bad, and The Ugly. *2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14)* (pp. 1–7). San Diego, CA: USENIX Association. Retrieved from https://www.usenix.org/conference/3gse14/summit-program/presentation/vigna

Švábenský, V., & Vykopal, J. (2018a, 02). Challenges Arising from Prerequisite Testing in Cybersecurity Games. *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (pp. 56–61). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3159450.3159454

Švábenský, V., & Vykopal, J. (2018b, 10). Gathering Insights from Teenagers' Hacking Experience with Authentic Cybersecurity Tools. *Proceedings of the 48th IEEE Frontiers in Education Conference* (pp. 1–4). New York, NY, USA: IEEE. Retrieved from https://doi.org/10.1109/FIE.2018.8658840

Švábenský, V., Vykopal, J., Čeleda, P. (2020, 03). What Are Cybersecurity Education Papers About? A Systematic Literature Review of SIGCSE and ITiCSE Conferences. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (pp. 2–8). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3328778.3366816

Švábenský, V., Vykopal, J., Čeleda, P., Kraus, L. (2022). *Dataset: Applications of Educational Data Mining and Learning Analytics on Data From Cybersecurity Training.* Zenodo. Retrieved from https://doi.org/10.5281/zenodo.6573117

Švábenský, V., Čeleda, P., Vykopal, J., Brišáková, S. (2020, 12). Cybersecurity Knowledge and Skills Taught in Capture the Flag Challenges. *Elsevier Computers & Security*, *102*(102154). Retrieved from https://doi.org/10.1016/j.cose.2020.1021548

Vykopal, J., & Barták, M. (2016). On the Design of Security Games: From Frustrating to Engaging Learning. *2016 USENIX Workshop on Advances in Security Education (ASE 16)* (pp. 1–8). Berkeley, CA, USA: USENIX Association. Retrieved from https://www.usenix.org/conference/ase16/workshop-program/presentation/vykopal

Vykopal, J., Vizvary, M., Oslejsek, R., Celeda, P., Tovarnak, D. (2017). Lessons learned from complex hands-on defence exercises in a cyber range. *2017 IEEE Frontiers in Education Conference (FIE)* (pp. 1–8). New York, NY, USA: IEEE. Retrieved from https://doi.org/10.1109/FIE.2017.8190713

Weiss, R., Locasto, M.E., Mache, J. (2016). A Reflective Approach to Assessing Student Performance in Cybersecurity Exercises. *Proceedings of the 47th ACM Technical Symposium on Computing Science Education* (pp. 597–602). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2839509.2844646

Weiss, R., Turbak, F., Mache, J., Locasto, M.E. (2017). Cybersecurity Education and Assessment in EDURange. *IEEE Security & Privacy*, *15*(3), 90–95. Retrieved from https://doi.org/10.1109/MSP.2017.54

Yamin, M.M., Katt, B., Gkioulos, V. (2020). Cyber ranges and security testbeds: Scenarios, functions, tools and architecture. *Computers & Security*, *88*, 101636. Retrieved from https://doi.org/10.1016/j.cose.2019.101636

Yett, B., Snyder, C., Zhang, N., Hutchins, N., Mishra, S., Biswas, G. (2020). Using log and discourse analysis to improve understanding of collaborative programming. *Proceedings of the 28th International Conference on Computers in Education* (pp. 1–10). Retrieved from https://apsce.net/icce/icce2020/proceedings/paper_158.pdf

Zeng, Z., Deng, Y., Hsiao, I., Huang, D., Chung, C.-J. (2018, 10). Improving student learning performance in a virtual hands-on lab system in cybersecurity education. *2018 IEEE Frontiers in Education Conference (FIE)* (p. 1–5). New York, NY, USA: IEEE. Retrieved from https://doi.org/10.1109/FIE.2018.8658855

Zurita, G., Shukla, A.K., Pino, J.A., Merigó, J.M., Lobos-Ossandón, V., Muhuri, P.K. (2020). A bibliometric overview of the journal of network and computer applications between 1997 and 2019. *Journal of Network and Computer Applications*, *165*, 102695. Retrieved from https://doi.org/10.1016/j.jnca.2020.102695