Research and Practice of Delivering Tabletop Exercises

Jan Vykopal
Masaryk University
Faculty of Informatics
Brno, Czech Republic
vykopal@fi.muni.cz

Pavel Čeleda
Masaryk University
Faculty of Informatics
Brno, Czech Republic
celeda@fi.muni.cz

Valdemar Švábenský
Masaryk University
Faculty of Informatics
Brno, Czech Republic
valdemar@mail.muni.cz

Martin Hofbauer
Masaryk University
Faculty of Informatics
Brno, Czech Republic
hofbauer@fi.muni.cz

Martin Horák
Masaryk University
Faculty of Informatics
Brno, Czech Republic
horak.martin@fi.muni.cz

ABSTRACT

Tabletop exercises are used to train personnel in the efficient mitigation and resolution of incidents. They are applied in practice to support the preparedness of organizations and to highlight inefficient processes. Since tabletop exercises train competencies required in the workplace, they have been introduced into computing courses at universities as an innovation, especially within cybersecurity curricula. To help computing educators adopt this innovative method, we survey academic publications that deal with tabletop exercises. From 140 papers we identified and examined, we selected 14 papers for a detailed review. The results show that the existing research deals predominantly with exercises that follow a linear format and exercises that do not systematically collect data about trainees' learning. Computing education researchers can investigate novel approaches to instruction and assessment in the context of tabletop exercises to maximize the impact of this teaching method. Due to the relatively low number of published papers, the potential for future research is immense. Our review provides researchers, tool developers, and educators with an orientation in the area, a synthesis of trends, and implications for further work.

CCS CONCEPTS

• General and reference → Surveys and overviews; **• Social and professional topics → Computing education**.

KEYWORDS

tabletop exercise, incident response, experiential learning, cybersecurity, hands-on training, systematic literature review

ACM Reference Format:
Jan Vykopal, Pavel Čeleda, Valdemar Švábenský, Martin Hofbauer, and Martin Horák. 2024. Research and Practice of Delivering Tabletop Exercises. In *Proceedings of the 2024 Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2024), July 8–10, 2024, Milan, Italy*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3649217.3653642

ITiCSE 2024, July 8–10, 2024, Milan, Italy

1 INTRODUCTION

The Computing Curricula 2020 report (CC2020) [17] responds to the need for graduates who are effective in their work roles and tasks. It promotes using *competencies* instead of knowledge to describe computing curricula. Competency augments knowledge with its skilled application motivated by the purpose of accomplishing a task. According to CC2020, the knowledge of subject matter and related skills are as important as analytical and critical thinking, collaboration and teamwork, and dispositions such as being responsible and flexible or having self-confidence and self-control.

Computing educators research and use various forms of experiential learning to prepare students for their future careers, such as simulations, team projects, or industry internships. We highlight *tabletop exercises* (TTX) as an efficient method for gaining competencies relevant to work roles and tasks centered around analysis, decision making, and communication with various parties. In particular, these exercises are relevant for cybersecurity and IT governance courses [12], addressing incident response methodologies outlined in frameworks like COBIT (Control OBjectives for Information and related Technology) and ITIL (Information Technology Infrastructure Library). While tabletop exercises are common in professional settings, they have not been widely used in computing courses. We believe that these exercises have a great potential for innovating teaching practice in higher education institutions.

This paper examines the development and state of the art of TTX as presented at various academic venues. The core contribution is a systematic review of research papers to understand state of the art. Our work is useful to various target groups. For educators, it shows examples of approaches and exercises and presents practical recommendations. For researchers, it provides an overview of methods used in exercises and implications for further research. Finally, developers of educational tools can be inspired by the existing tools and implications for future work.

2 BACKGROUND

A tabletop exercise is a form of a teaching activity aimed at training teams in responding to crisis situations [19]. It involves simulating a crisis within the context of business operations in an organization, such as a ransomware attack on the company infrastructure or exfiltration of sensitive information. The team members (exercise participants) assume different roles in the organization, such as chief security officer or cybersecurity incident responder [3]. During the exercise, they discuss how to effectively respond to the crises while adhering to processes and regulations. Instructors facilitate these discussions and provide a debriefing after the TTX. The exercise usually lasts a few hours or days at most.

TTXs are driven by *injects*, pre-scripted messages, which can take the form of an e-mail or a news article, provided to trainees during the exercise. The purpose of injects is to advance the exercise and stimulate further actions and discussions. For instance, injects can notify teams of a data breach in their organization, requiring them to respond appropriately [19].

TTXs focus on communication, coordination, and collaboration [3], not particular technical skills as it is the case during hands-on training in an emulated IT environment (such as a cybersecurity lab [20] or cyber range [50]). The nature of TTXs allows to conduct them using pen and paper or simple online office applications (such as Microsoft Forms), making TTXs relatively cost-effective. However, assessing the trainees is not automated – it requires instructors' manual effort, which is highly time-consuming. It may take days or weeks until the trainees receive feedback from instructors, which lowers the effectiveness of the exercise.

TTXs share certain traits with some other forms of active learning [43] but also have unique characteristics. For instance, course projects in software development or Process oriented guided inquiry learning (POGIL) [27] also involve student teams that work together. In contrast, TTXs do not feature a clearly specified and structured assignment. Moreover, student teams in a TTX are not guided by instructors but decide on their own about the current priorities and tasks to complete. Last, TTXs intentionally overwhelm trainees with numerous pieces of information and inputs to simulate a stressful emergency situation.

TTXs in the cybersecurity domain can be conducted either independently or as a part of complex exercises involving technical skills (denoted as cyber defense exercises or red vs. blue team exercises). Locked Shields [34] and Cyber Europe [16] are exercises combining TTXs with technical training. Both are centered around a background story resembling a recent real crisis or attack campaign [49, 52]. Participants representing one organization or country are assigned different roles and divided into teams. While some teams exercise mainly technical skills in an emulated IT environment, others are engaged in decision-making processes, standard operational procedures, or communication in a local and international context. Locked Shields is the largest global defense exercise organized by the NATO Cooperative Cyber Defence Centre of Excellence in Tallinn, Estonia, since 2010. Cyber Europe is a series of pan-European exercises developed for IT security, business continuity, and crisis management teams by European Union Agency for Cybersecurity (ENISA), since 2010.

Diverse commercial solutions are available to address various aspects of emergency preparation. Some vendors, such as PreparedEx [40] or Emergency Solutions International [21], provide comprehensive TTXs for crises, training emergency responders, and highlighting inefficiencies in processes and inadequate capabilities. Numerous commercial services (e.g., [7, 8, 30, 41]) focus on validating incident response and business continuity plans in IT operations or cybersecurity and are often part of broader services portfolios. Additionally, there are platform-as-a-service solutions (namely [4, 6, 10, 53]), all focused on crisis management in general. Furthermore, specialized platforms, such as Cyber Crisis Simulator [28] or open-source OpenEx [15], are designed specifically for IT and cybersecurity crisis management.

National or international authorities, such as NIST [19], FEMA [14], ENISA [12], or ISO [22] published standards and guides on complex exercises, which also include TTXs or can be applied to them.

To conclude, a TTX is an established training method used in practice, yet mostly outside university settings. Our review maps the state of the art based on academic publications on this topic.

3 PREVIOUS LITERATURE REVIEWS

In the academic literature, there is one review of TTXs in the cybersecurity domain and three papers in the subject area of healthcare.

Angafor et al. [2] reviewed academic and commercial product literature on TTXs used for training cybersecurity incident response teams. The scope of our review is wider; we searched for papers in broad subject areas of computer science and engineering. Also, our review captures recent research and trends since 2020 when the other review was published. The overlap of this and their study is only two papers (namely P4 and P7, see details in Table 1).

Mahdi et al. [31] reviewed disaster preparation exercises conducted by academic healthcare institutions. Based on the reviewed literature, the authors concluded that TTXs are the easiest to organize, conduct, and evaluate, while also useful in evaluating emergency response protocols and their subsequent improvement.

Evans [13] surveyed healthcare literature for using TTXs in nursing education. The opportunity to identify knowledge gaps, as well as knowledge gain, was reported. The author also proposed a list of considerations for exercise development.

Finally, Frégeau et al. [18] published a scoping protocol for a review that maps the uses of TTXs in healthcare. However, the review itself has not been published yet.

4 METHOD OF CONDUCTING THE REVIEW

We follow the guidelines for a systematic literature review (SLR) [24] and a systematic mapping study [38, 39]. This section presents the SLR protocol, which specifies the research questions, search process (see Figure 1), and criteria for including the discovered papers.

4.1 Research Questions

Our literature review examines five research questions:

(1) *What formats of tabletop exercises are used?*
Namely how the exercises are prepared, delivered and what tools are applied.

(2) *Who are the participants of the exercises?*
What are the target groups of the exercises (trainees) and by whom are they organized (instructors, designers)?

(3) *How are the exercises developed, assessed, and evaluated?*
We will examine methods for exercise preparation, assessment of trainees, and evaluation of the exercise itself.

(4) *How are the research results applied in practice?*
Do the publications provide any supplementary materials or artifacts for other educators?

(5) *What are the future research directions and challenges?*
What is the ongoing and long-term goal?

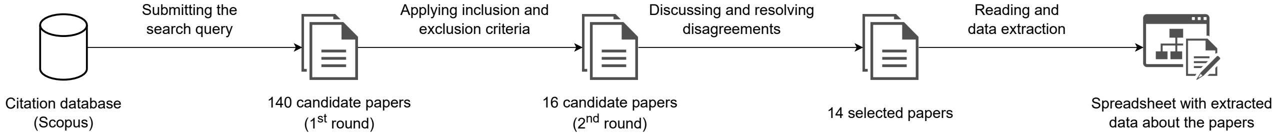
Submitting the search query
Applying inclusion and exclusion criteria
Discussing and resolving disagreements
Reading and data extraction
Citation database (Scopus)
140 candidate papers (1st round)
16 candidate papers (2nd round)
14 selected papers
Spreadsheet with extracted data about the papers

Figure 1: Steps of the systematic literature review and the number of processed papers.

```
TITLE-ABS-KEY( "table* exercise" OR "communication exercise" ) AND ( LIMIT-TO ( SUBJAREA, "COMP" ) OR LIMIT-TO ( SUBJAREA, "ENGI" ) )
```

Figure 2: The query for the automated paper search for papers in Scopus [11].

4.2 Paper Search and Automated Filtering

We used the Scopus citation database of peer-reviewed literature [11]. Scopus indexes a representative portion of the databases of individual publishers, including ACM Digital Library, IEEExplore, SpringerLink, or Elsevier ScienceDirect. We did not use Google Scholar since it additionally indexes many non-peer-reviewed papers.

When constructing the search query, we focused on the subject area of computer science and engineering and used "tabletop exercise" as the primary keyword. We also sought not to miss potential alternative names for tabletop exercises, such as "simulation" or "communication exercises". After several pilot searches, we concluded with the query presented in Figure 2. The asterisk is a wildcard to match both adjectives "table top" and "table-top". However, the wildcard is not used for the stem of the word "communication" because the pilot search yielded numerous false positives. Similarly, we omitted "simulation" matching a huge number of papers unrelated to education and training. The search was case-insensitive.

We started the search on May 11, 2023, and then 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 July 19, 2023. In total, we found 140 candidate papers.

4.3 Inclusion and Exclusion Criteria

After multiple pilot tests, we set ten criteria.

(1) Papers must deal with TTXs in IT or operational technology (OT) operations or security.
(2) Papers describing exercises and tools in other domains, such as transportation or utilities, without considering cybersecurity or IT/OT operations aspects, are excluded.
(3) We include papers written in English with *full text available*[1]. We do *not set any page limit* to include short papers, which report ongoing and future work at the time of their publication.
(4) Other types of documents yielded by the automated search, such as conference reviews or records, are excluded.
(5) Papers must describe *exercising or supporting the exercise of a complex process or standard operation procedures in a safe, simulated environment.* For instance, this includes experience reports with lessons learned or descriptions of tools supporting the preparation, execution, and analysis of TTXs.
(6) Papers that describe only exercising essential communication and writing skills (e.g., technical writing, proposal paper, presentation and its delivery) are excluded.
(7) Papers must report on an exercise involving *teams or groups, not only individuals.*
(8) The paper must support an *educational goal*, for example, to train or assess participants, or help instructors conduct the exercise, or understand the learning processes.
(9) We exclude papers that use TTXs solely as an evaluation tool in non-educational settings, such as for testing of software or IT systems.
(10) *Generic methodologies* applied or applicable to TTXs in IT and OT operations and security are included.

Two authors of this paper each screened all 140 candidate papers and applied these criteria independently. The authors followed a simple algorithm for screening the paper content [46]:

```
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
```

4.4 Selecting Papers for Review

After both authors finished their reading, they compared candidate papers they identified after applying the inclusion and exclusion criteria. If both authors decided to include the paper, it was selected for review. If both decided not to include the paper, it was excluded. In case of a disagreement, the authors discussed their views and agreed on the final decision.

In the first round, the authors agreed that out of 140 candidate papers, nine (6.4%) were fitting the selection criteria. The authors had opposite opinions on seven (5%) papers. Out of the 16 papers, which passed the first round, the authors selected 14 papers and rejected two papers in the second selection round. Their inter-rater agreement [26] measured by Krippendorff's α for nominal data was 0.69, which is substantial. The coefficient was calculated using the Python NLTK module [35].

5 RESULTS

This section presents data extracted from reading the full texts of the 14 papers to answer the research questions posed in Section 4.1.

[1] We consider a paper to be available if we can access it using tens of licensed electronic resources our institution has access to, or if the full text is freely available online.

Table 1 summarizes the goals and types of the reviewed papers. The summary of the papers is provided in Section 6. Further, we refer to the selected papers using arbitrary numbered identifiers P_x.

5.1 RQ1: What Formats of Exercises Are Used?

From seven papers (namely P1, P2, P4, P7, P12, P13, P14) out of eight that contain enough information about the exercise format, we see that the exercises are designed as a series of injects (events, problems, or situations, see Section 2). These injects form a scenario that is unknown to the exercise participants beforehand. The injects are provided by exercise facilitators to participants who discuss appropriate actions, processes, or best practices within their team or with all participants.

The eighth paper (P11) reports two kinds of exercises that extend this common format. One TTX uses cards that drive the exercise itself (attack cards) or stimulate participants to think about several options (action cards, situation awareness cards, and information-sharing cards). Another TTX in P11 tasked participants to create a scheme depicting information and workflows within an organization affected by an incident.

Only four papers mention the use of any software tool during the exercise. P7 presents a web application for delivering a TTX involving various roles according to a scenario defined by the facilitator. Participants use a graphical interface to respond to presented injects. They can submit a short description of how they would solve the inject, inform other roles about it, or delegate the resolution to other roles. P11 mentions an exercise where participants use software to interact with virtual participants. Participants choose a response to presented injects from a limited set of predefined actions. P12 reports a remote exercise facilitated through Microsoft Teams web conferencing application. P1, P6, and P14 use a software tool before or after the actual exercise; see Section 5.3.

The exercises last from several hours (P2, P13, P14) through one day (P4, P7, P8, P9, P11, P12) to a few days (P6).

5.2 RQ2: Who Are the Exercise Participants?

Trainees come from diverse sectors. The most frequent were critical infrastructure organizations, such as energy distribution operators (P4), a water management centre (P6), industrial control systems stakeholders (P8, P11), or oil and gas suppliers (P9, P13). Two exercises were carried out for university students (P2, P7). One exercise was conducted for a large law enforcement organization (P12). The number of trainees ranged from 20 to 108.

Exercises were designed by representatives of national or transnational authorities (P4, P9) or academic staff (P2, P8, P11, P12). The type of the organizing entity determines the target group and its diversity (employees of one organization, multiple organizations in a single country or multiple organizations from multiple countries).

5.3 RQ3: How Are the Exercises Developed, Assessed, and Evaluated?

Development. There is no prevailing trend in the process of exercise preparation in the reviewed papers, even though guidelines [12, 14, 19] and standards [22] had already been published.

The exercises in P4 and P7 were designed and conducted using guidelines from NIST [19] and ENISA [12], respectively.

P1 introduces a web-based collaborative tool for designers of cybersecurity TTXs, enabling scenario creation based on TTX objectives following a national guideline for exercise development [14]. It supports various user roles like Scenario Designer, Subject Matter Expert, or Observer, in producing a detailed list of scenario events.

P5 studies the compliance of three guidelines from the European authorities for designing and conducting exercises with the international standard ISO 22398 [22].

P13 investigates the characteristics of a realistic and expedient scenario in the field of industrial control systems. The paper lists 21 criteria and eight exercise topics based on these scenarios.

P3 argues that current tabletop exercises do not accurately reflect the reality of uncertain and complex problems that do not have obvious solutions. It presents three design ideas for designing more efficient exercises. First, the focus should be on unsolved problems that the participants themselves come up with. Second, problems should be tamed during the exercise by the participants instead of during the planning phase by the designer. Third, the participants should use existing plans and experience from previous emergencies to resolve the problems in collaboration.

P14 employs machine learning to generate scenarios of cybersecurity exercises. First, a corpus of publicly available articles about cybersecurity incidents is created and annotated to detect threat actors, incidents, and victims. Then, an exercise designer provides inputs such as a keyword or a sector for generating a graph of an incident for a created exercise. The graph is then enhanced using information mined from existing cybersecurity taxonomies. Finally, the graph is transformed to text using GPT-2.

Assessment. The assessment of exercise participants (trainees and/or facilitators) is addressed only in three papers.

P6 proposes a method and a tool for structured assessment of trainees. First, observable behavior in exercise subgoals must be defined in the exercise preparation. During the exercise, human observers record five performance aspects (timeliness, accuracy, relevance, completeness, and cost-effectiveness) in four phases of the OODA loop (Observe, Orient, Decide, Act) [42], and the tool produces a score capturing trainees' performance.

Exercises reported in P4 include unstructured assessment of trainees' actions by the observers and facilitators after the exercise.

P8 studies errors made by exercise facilitators during their interactions with groups of trainees. P8 recognizes the error of *commission* and the error of *omission*, which can negatively affect the progress of a trainee group in the exercise.

Evaluation. Six papers addressed the evaluation of the exercise itself. None reported the use of specific qualitative or quantitative research methods.

In P4 and P11, the evaluation is conducted as a reflection of trainees and a discussion with the exercise designers and facilitators after the exercise.

In the exercise described in P2, teams of trainees compete against themselves in opposing roles of attackers and defenders. The feedback session is then used to uncover the goals and motivations of the other role. Also, the instructors ask for immediate feedback on the exercise format.

P7 presents summary counts of trainees' actions in the exercise platform and messages sent during the exercise.

Table 1: Overview of the 14 reviewed papers ordered by the year of publication. Paper Type: SW = software tool, E = exercise instance (run of a particular exercise), F = exercise format (method for conducting the exercise), O = other topic (see Goal).

Paper ID	Year	Paper Type	Goal of the Paper
P1 [33]	2009	SW, F	Describe a tool for planning complex functional and tabletop exercises
P2 [37]	2014	F, E	Describe a format of lightweight exercises
P3 [9]	2014	O	Propose design ideas helping to incorporate more problems having no obvious solution into TTXs
P4 [29]	2015	E	Study challenges when performing TTXs
P5 [32]	2015	O	Review of Hermes, MSB and ENISA exercise methodologies w.r.t. the ISO standard 22398
P6 [23]	2015	O, E	Introduce a method for trainee assessment
P7 [5]	2017	SW, F, E	Present a web-based environment for conducting TTXs
P8 [45]	2019	E, O	Study errors made by TTX facilitators when interacting with trainees
P9 [25]	2020	E	Report takeaways from a particular exercise
P10 [1]	2020	O	Investigate the current cybersecurity skills gap and how TTXs can fill it
P11 [36]	2022	F, E	Describe three exercises developed by the authors
P12 [3]	2023	O, E	Report experience from a virtual incident response tabletop exercise
P13 [44]	2023	O, F	Investigate the characteristics of a realistic and expedient exercise scenario
P14 [51]	2023	SW, F	Apply machine learning to unstructured information sources to generate exercise content

P9 reports the evaluation that has two phases. Firstly, the trainees shared comments about what they learned at the end of the exercise. Afterward, experts invited by the exercise organizers surveyed the trainees and evaluated the outputs produced during the exercise.

P12 discusses responses from a trainee survey, which relates the exercise content to incident response and disaster recovery plans of a particular organization where the exercise was held.

5.4 RQ4: How Are the Results Applied?

No reviewed paper refers to supplementary materials (such as software tool implementation or exercise scenario) that other educators can directly use. In particular, P1, P6, P7, and P14 describe software tools, but the papers do not refer to any materials, such as software repositories or websites presenting the tools. P13 only outlines eight scenarios as an inspiration for creating a new exercise.

A few papers distill recommendations and lessons learned from TTX preparation or delivery applicable to other exercises. P2 lists steps describing the preparation of a lightweight TTX and the most common problems encountered. P3 provides three design ideas for more realistic and efficient exercises (see Section 5.3). P4 states recommendations for conducting exercises.

Finally, P5 shows that the methodology by ENISA [12] is the most compliant with the ISO 22398 standard out of the three reviewed methodologies, yet not entirely.

5.5 RQ5: What Are the Future Directions?

The reviewed papers reported diverse future work. P2 suggests developing more detailed instructions, creating mock scenarios, and formalizing and publishing student feedback. P3 advises the development of an online tool for running a TTX, evaluating the proposed design ideas, and studying how TTXs are used in incident recovery. P4 recommends studying best practices and challenges of organizations conducting TTXs regularly and studying real-life incident responses to design more useful future exercises. P6, which deals with assessment, plans to replace binary scoring with a four-point scale in the presented method. Since P7 piloted the exercise with university students, future work includes conducting the exercise with professionals working in the target industry. In addition, it propounds the development of an assessment of trainee stress levels, which can then be used to adapt the exercise progression dynamically. P8 mentions sharing the instances of errors among facilitators. While P9 simply recommends conducting more exercises, P11 suggests developing a method to demonstrate the risks due to failure to act. P12 plans to improve the presented exercise using lessons learned from its first run and develop a training program for staff at all organizational levels. P13 would like to test the proposed criteria for realistic and expedient scenarios in practice when deploying the proposed scenarios. Finally, P14 lists several concrete directions to improve the tool for generating exercise content.

6 DISCUSSION

This section provides takeaways from the review of the selected papers. It helps the reader understand the greater themes of the results presented in Section 5. The review limitations are also discussed.

6.1 Summary of the Observed Trends

The topics in the identified papers differ widely. Most papers are reports from runs of a particular exercise or describe an exercise format. The typical TTX format is based on pieces of information gradually provided to the trainees by exercise facilitators. The trainees discuss and respond to the information in teams. The exercises are usually held for tens of trainees on site for several hours or days, with little support from dedicated software tools.

The explicit learning phase usually happens at the end or after the exercise, when trainees reflect on their decisions with other participants or when the previously unknown parts of the scenario are disclosed to them. Other types of assessment or feedback are rare because the complexity and labor of the manual preparation of the exercises hinder instructors from focusing on trainee assessment and exercise evaluation. Out of only three papers (P4, P6, and P8) that addressed assessment, only one (P6) suggested a method that goes further than unstructured assessment by the observers and facilitators. Similarly, the evaluation of the exercise itself is also underdeveloped. No paper uses advanced feedback methods but trainees' reflections, discussions, or surveys.

The reviewed papers did not provide many actionable artifacts or supplementary materials (such as software supporting exercise preparation and delivery, exercise playbooks, scenarios, or checklists) for others considering conducting their own exercise. The absence of publicly available materials prevents educators from adopting the concept of tabletop exercises as a teaching method.

The current practice of conducting TTXs relies on manual preparation of exercise content (namely injects that form exercise scenarios), with no automation and limited future reusability. We believe this practice can be improved by leveraging large language models for semi-automated content generation (P14) or employing dedicated software for TTX preparation and delivery (see Section 5.1).

Finally, future directions in the papers often mention evaluating proposed methods by field tests or running the exercise for more trainees or different target groups. These plans indicate that the described methods and exercises should be developed further and provide new research opportunities in computing education.

6.2 Limitations

This review is limited by narrowing its scope by only using the Scopus citation database of the peer-reviewed literature. This decision was explained in Section 4.2. Also, when using Scopus, we did not have to deal with duplicate records, which may occur when using multiple other databases.

The second limitation is keyword choice. Using other keywords may have led us to find different papers fitting the selection criteria. However, pilot searches with related terms, such as "crisis exercise" returned an excessive number of unrelated papers (see Section 4.2).

Finally, the data extraction was done manually, which may lead to misinterpretation of the presented information. To mitigate this risk, we performed cross-author discussion and validation.

7 CONCLUSIONS AND FUTURE DIRECTIONS

While tabletop exercises have been conducted for over a decade, often by governmental and military agencies, the number of published papers has grown only in recent years. The organizing entities have developed their guidelines and summarized the best practices, but there is no widely used terminology, textbooks, or tools for the exercise development, delivery, or evaluation. This situation is one of the causes why TTXs are not yet widely used in computing education, even though they can help develop competencies and dispositions relevant to students' future careers as IT professionals.

To the best of our knowledge, this paper is the first systematic review of TTXs in the context of IT operations and security. We examined 140 papers and reviewed 14 of them in detail. The structured information extracted from the identified papers is published as supplementary material [48], which also includes a review of 16 practical implementations that did not fit this paper's page limit.

This review implies several possible directions for future research and practice. First, novel formats of the exercises can be proposed to make them more realistic but still lightweight on resources and effort required for their development and delivery. For instance, the current discussion-based approach ("What would your team/organization do when a certain event happened?") could be innovated by letting the trainees perform the actual simulated actions (such as asking another team/external party to do something) and observe whether they do what was expected by the exercise scenario based on training objectives. Another example is an investigation of opportunities and limits of remote or hybrid exercise that would enable more trainees to participate and ease traveling and logistics.

The next direction is to start using dedicated software tools in the TTX preparation, delivery, and evaluation, such as INJECT Exercise Platform [47]. Automated tools can substantially increase exercise scalability and lower the preparation effort. Not only can they enable the participation of more trainees (regardless of whether the TTX is delivered onsite or remotely), but they also reduce the workload of exercise designers and facilitators. Moreover, researchers may explore employing machine learning for trainee assessment or leverage large language models to generate exercise content.

Next, developers can provide a dedicated platform for designing and conducting the exercise, which would assist the facilitators or even automate some of their tasks (such as playing the role of some simulated actors in the exercise). Moving the exercise from pen and paper or online forms and documents to the dedicated platform would enable the collection of data about trainees' interactions during the exercise and further analysis for the assessment of trainees and evaluation of the exercise. Finally, the exercise designers would strongly benefit from releasing and sharing tangible outputs, such as software tools, scenarios, or checklists.

ACKNOWLEDGMENTS

This research was supported by the Open Calls for Security Research 2023–2029 (OPSEC) program granted by the Ministry of the Interior of the Czech Republic under No. VK01030007 – *Intelligent Tools for Planning, Conducting, and Evaluating Tabletop Exercises.*

REFERENCES

[1] Giddeon N. Angafor, Iryna Yevseyeva, and Ying He. 2020. Bridging the Cyber Security Skills Gap: Using Tabletop Exercises to Solve the CSSG Crisis. In *Serious Games*. Springer International Publishing, Cham, 117–131. https://doi.org/10.1007/978-3-030-61814-8_10

[2] Giddeon N. Angafor, Iryna Yevseyeva, and Ying He. 2020. Game-based learning: A review of tabletop exercises for cybersecurity incident response training. *Security and Privacy* 3, 6 (2020), e126. https://doi.org/10.1002/spy2.126

[3] Giddeon N. Angafor, Iryna Yevseyeva, and Leandros Maglaras. 2023. Scenario-based incident response training: lessons learnt from conducting an experiential learning virtual incident response tabletop exercise. *Information & Computer Security* 31, 4 (2023). https://doi.org/10.1108/ICS-05-2022-0085

[4] Avalias. 2023. Avalanche TTX. Retrieved January 17, 2024 from https://www.avalias.com/products/avalanche-ttx

[5] Agnė Brilingaitė, Linas Bukauskas, Virgilijus Krinickij, and Eduardas Kutka. 2017. Environment for Cybersecurity Tabletop Exercises. In *11th European Conference on Games Based Learning*. Graz, Austria, 47–55. https://www.researchgate.net/publication/320244434_Environment_for_Cybersecurity_Tabletop_Exercises

[6] Cinten. 2023. Crisis Management. Retrieved January 17, 2024 from https://www.cinten.com/crisis-management-digital-simulations-performance-analytics

[7] Cyber Security Operations Consulting. 2023. *Cybersecurity TTX and Incident Response*. CyberSecOp. Retrieved January 17, 2024 from https://cybersecop.com/tabletop-exercise/cybersecurity-tabletop-exercise-services

[8] CrowdStrike. 2023. *CrowdStrike TTX*. CrowdStrike. Retrieved January 17, 2024 from https://www.crowdstrike.com/services/prepare/tabletop-exercise/

[9] Svante Edzén. 2014. Table-Top Exercises for Emergency Management: Tame Solutions for Wicked Problems. In *2014 47th Hawaii International Conference on System Sciences*. IEEE, USA, 1978–1985. https://doi.org/10.1109/HICSS.2014.250

[10] eedo inc. 2023. *Crisis & Preparedness Cooperation Exercise Simulator*. eedo inc. Retrieved January 17, 2024 from https://eee.do/crisis-and-preparedness-cooperation-exercise-simulation-system/

[11] Elsevier. 2024. Scopus. Retrieved January 17, 2024. https://www.scopus.com

[12] ENISA. 2009. *Good Practice Guide on National Exercises*. Technical Report. European Union Agency for Cybersecurity. https://www.enisa.europa.eu/publications/national-exercise-good-practice-guide

[13] Cathleen A. Evans. 2019. Tabletop exercises in the nursing classroom: An introduction for nurse educators. *Nursing Forum* 54, 4 (2019), 669–674. https://doi.org/10.1111/nuf.12394

[14] Federal Emergency Management Agency. 2020. *Homeland Security Exercise and Evaluation Program (HSEEP)*. Technical Report. Federal Emergency Management Agency. https://www.fema.gov/sites/default/files/2020-04/Homeland-Security-Exercise-and-Evaluation-Program-Doctrine-2020-Revision-2-2-25.pdf

[15] Filigran. 2024. *OpenEx Platform*. Filigran. Retrieved January 17, 2024 from https://github.com/OpenEx-Platform/openex

[16] ENISA (European Union Agency for Cybersecurity). 2023. Cyber Europe. Online, accessed January 17, 2024. https://www.enisa.europa.eu/topics/training-and-exercises/cyber-exercises/cyber-europe-programme

[17] CC2020 Task Force. 2020. *Computing Curricula 2020: Paradigms for Global Computing Education*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3467967

[18] Amélie Frégeau, Alexis Cournoyer, Marc-André Maheu-Cadotte, Massimiliano Iseppon, Nathalie Soucy, Julie St-Cyr Bourque, Sylvie Cossette, Véronique Castonguay, and Richard Fleet. 2020. Use of tabletop exercises for healthcare education: a scoping review protocol. *BMJ Open* 10, 1 (2020). https://doi.org/10.1136/bmjopen-2019-032662

[19] Tim Grance, Tamara Nolan, Kristin Burke, Rich M. Dudley, Gregory C. White, and Travis Good. 2006. *Guide to Test, Training, and Exercise Programs for IT Plans and Capabilities: Recommendations of the National Institute of Standards and Technology*. Technical Report. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-84

[20] Philip Huff, Sandra Leiterman, and Jan P. Springer. 2023. Cyber Arena: An Open-Source Solution for Scalable Cybersecurity Labs in the Cloud. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2023)*. Association for Computing Machinery, New York, NY, USA, 221–227. https://doi.org/10.1145/3545945.3569828

[21] Emergency Solutions International Inc. 2023. *ESI Exercises*. ESI. Retrieved January 17, 2024 from https://esintl.ca/exercise-and-evaluation/

[22] ISO. 2013. *Societal security — Guidelines for exercises*. Standard. International Organization for Standardization, Geneva, CH.

[23] Astrid Janssen and Hanneke Vreugdenhil. 2015. Objective oriented exercise evaluation with TARCK-it. In *12th Proceedings of the International Conference on Information Systems for Crisis Response and Management, Krystiansand, Norway, May 24-27, 2015*. ISCRAM Association, Kristiansand, Norway. http://idl.iscram.org/files/astridjanssen/2015/1223_AstridJanssen+HannekeVreugdenhil2015.pdf

[24] Barbara Kitchenham and Stuart Charters. 2007. *Guidelines for performing Systematic Literature Reviews in Software Engineering*. Technical Report. EBSE.

[25] Vytis Kopustinskas, Bogdan Vamanu, Marcelo Masera, Rimantas Šikas, Julia Vainio, Romualdas Petkevičius, and Lawrence Walzer. 2020. Tabletop Exercise as a Tool to Foster Resilience in the Energy Sector: Lessons Learned in the Baltic States. In *Proceedings of the 30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference*. Research Publishing Services, Italy, 255–262. https://doi.org/10.3850/978-981-14-8593-0_4019-cd

[26] Klaus Krippendorff. 2004. Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research* 30, 3 (2004), 411–433. https://doi.org/10.1111/j.1468-2958.2004.tb00738.x

[27] Clifton Kussmaul. 2012. Process Oriented Guided Inquiry Learning (POGIL) for Computer Science. In *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education (SIGCSE '12)*. Association for Computing Machinery, New York, NY, USA, 373–378. https://doi.org/10.1145/2157136.2157246

[28] Immersive Labs. 2023. *Cyber Crisis Simulator*. Immersive Labs. Retrieved January 17, 2024 from https://www.immersivelabs.com/platform/cyber-crisis-simulator/

[29] Maria B. Line and Nils B. Moe. 2015. Understanding Collaborative Challenges in IT Security Preparedness Exercises. In *ICT Systems Security and Privacy Protection*. Springer, Cham, 311–324. https://doi.org/10.1007/978-3-319-18467-8_21

[30] Micro Minder Ltd. 2023. *Cybersecurity Tabletop Exercise Services*. MCS. Retrieved January 17, 2024 from https://www.microminder cs.com/cybersecuritytabletopexercise

[31] Syed S. Mahdi, Hafsa A. Jafri, Raheel Allana, Gopi Battineni, Mariam Khawaja, Syeda Sakina, Daniyal Agha, Kiran Rehman, and Francesco Amenta. 2023. Systematic review on the current state of disaster preparation Simulation Exercises (SimEx). *BMC Emergency Medicine* 23, 1 (2023), 52. https://doi.org/10.1186/s12873-023-00824-8

[32] Georgios Makrodimitris and Christos Douligeris. 2015. Towards a Successful Exercise Implementation – A Case Study of Exercise Methodologies. In *Human Aspects of Information Security, Privacy, and Trust*. Springer International Publishing, Cham, 207–218. https://doi.org/10.1007/978-3-319-20376-8_19

[33] Jim Marshall. 2009. The Cyber Scenario Modeling and Reporting Tool (CyberSMART). In *2009 Cybersecurity Applications & Technology Conference for Homeland Security*, Vol. 1. IEEE, USA, 305–309. https://doi.org/10.1109/CATCH.2009.46

[34] NATO Cooperative Cyber Defence Centre of Excellence. 2023. Locked Shields. Online, accessed January 17, 2024. https://ccdcoe.org/exercises/locked-shields

[35] Natural Language Toolkit (NLTK) Project. 2023. Source code for `nltk.metrics.agreement`. Online, accessed January 17, 2024. http://www.nltk.org/_modules/nltk/metrics/agreement.html

[36] Yuitaka Ota, Erika Mizuno, Tomomi Aoyama, Yoshihiro Hashimoto, Ichiro Koshijima, Haruna Asai, and Shiho Taniuchi. 2022. Designing Framework for Tabletop Exercise to Promote Resilience Against Cyber Attacks. In *14th International Symposium on Process Systems Engineering*. Elsevier, Japan, 1471–1476. https://doi.org/10.1016/B978-0-323-85159-6.50245-1

[37] Rain Ottis. 2014. Light Weight Tabletop Exercise for Cybersecurity Education. *Journal of Homeland Security and Emergency Management* 11 (12 2014), 579–592. Issue 4. https://doi.org/10.1515/jhsem-2014-0031

[38] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. 2008. Systematic Mapping Studies in Software Engineering. In *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering* (Italy) *(EASE'08)*. BCS Learning & Development Ltd., Swindon, UK, 68–77. https://doi.org/10.14236/ewic/EASE2008.8

[39] Kai Petersen, Sairam Vakkalanka, and Ludwik Kuzniarz. 2015. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology* 64 (2015), 1 – 18. https://doi.org/10.1016/j.infsof.2015.03.007

[40] LLC PreparedEx. 2023. *PreparedEx Services*. PreparedEx, LLC. Retrieved January 17, 2024 from https://preparedex.com/services/

[41] RedLegg. 2023. *RedLegg TTX*. GLW Specialty. Retrieved January 17, 2024 from https://www.redlegg.com/advisory-services/tabletop-exercise

[42] Richards, Chet. 2020. Boyd's OODA loop. *Necesse* 5 (2020), 142–165. https://hdl.handle.net/11250/2683228

[43] Kate Sanders, Jonas Boustedt, Anna Eckerdal, Robert McCartney, and Carol Zander. 2017. Folk Pedagogy: Nobody Doesn't Like Active Learning. In *Proceedings of the 2017 ACM Conference on International Computing Education Research* (Tacoma, Washington, USA) *(ICER '17)*. Association for Computing Machinery, New York, NY, USA, 145–154. https://doi.org/10.1145/3105726.3106192

[44] Andrea Skytterholm and Guro Hotvedt. 2023. Criteria for Realistic and Expedient Scenarios for Tabletop Exercises on Cyber Attacks Against Industrial Control Systems in the Petroleum Industry. In *Proceedings of the International Conference on Cybersecurity, Situational Awareness and Social Media*. Springer Nature, Singapore, 39–54. https://doi.org/10.1007/978-981-19-6414-5_3

[45] Shiho Taniuchi, Tomomi Aoyama, Haruna Asai, and Ichiro Koshijima. 2019. Training Cyber Security Exercise Facilitator: Behavior Modeling Based on Human Error. In *Advances in Human Factors in Cybersecurity*. Springer International Publishing, Cham, 138–148. https://doi.org/10.1007/978-3-319-94782-2_14

[46] Valdemar Švábenský, Jan Vykopal, Pavel Čeleda, and Lydia Kraus. 2022. Applications of educational data mining and learning analytics on data from cybersecurity training. *Education and Information Technologies* 27, 9 (2022), 12179–12212. https://doi.org/10.1007/s10639-022-11093-6

[47] Valdemar Švábenský, Jan Vykopal, Martin Horák, Martin Hofbauer, and Pavel Čeleda. 2024. From Paper to Platform: Evolution of a Novel Learning Environment for Tabletop Exercises. In *Proceedings of the 29th Conference on Innovation and Technology in Computer Science Education (ITiCSE '24)*. Association for Computing Machinery, New York, NY, USA, 7 pages. https://doi.org/10.1145/3649217.3653639

[48] Jan Vykopal, Pavel Čeleda, Valdemar Švábenský, Martin Hofbauer, and Martin Horák. 2024. Dataset: Research and Practice of Delivering Tabletop Exercises. https://gitlab.fi.muni.cz/inject/papers/2024-iticse-research-practice.

[49] Jan Vykopal, Martin Vizvary, Radek Oslejsek, Pavel Celeda, and Daniel Tovarnak. 2017. Lessons Learned From Complex Hands-on Defence Exercises in a Cyber Range. In *2017 IEEE Frontiers in Education Conference (FIE)*. IEEE, Indianapolis, USA, 1–8. https://doi.org/10.1109/FIE.2017.8190713

[50] Muhammad Mudassar Yamin, Basel Katt, and Vasileios Gkioulos. 2020. Cyber ranges and security testbeds: Scenarios, functions, tools and architecture. *Computers & Security* 88 (2020), 101636. https://doi.org/10.1016/j.cose.2019.101636

[51] Alexandros Zacharis and Constantinos Patsakis. 2023. AiCEF: an AI-assisted cyber exercise content generation framework using named entity recognition. *International Journal of Information Security* 22, 5 (2023), 1333–1354. https://doi.org/10.1007/s10207-023-00693-z

[52] Grethe Østby, Kieren N. Lovell, and Basel Katt. 2019. EXCON Teams in Cyber Security Training. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, Las Vegas, USA, 14–19. https://doi.org/10.1109/CSCI49370.2019.00010

[53] Conducttr ™. 2023. Crisis simulation platform. Retrieved January 17, 2024 from https://www.conducttr.com/