Evaluating Case Study and Action Research Reports: Real-world Research in Cybersecurity case study

Abstract: There is a growing number of scientific papers reporting on case stud- ies and action research published each year. Consequently, evaluating the quality of pilling up research reports is becoming increasingly challenging. Several approaches for evaluation of quality of the scientific outputs exist however they appear to be fairly time-consuming and/or adapted for other research designs. In this paper, we propose a reasonably light-weight structure-based approach for evaluating case study and action research reports (SAE-CSAR) based on eight key parts of a real-world research report: research question, case description, data collection, data analysis, eth- ical considerations, results, discussion and limitations. To evaluate the feasibility of the proposed approach, we conducted a systematic literature survey of papers reporting on real-world cybersecurity research. A total of N = 102 research papers were evalu- ated. Results suggest that SAE-CSAR is useful and relatively efficient, and may offer a thought-provoking insight into the studied field. Although there is a positive trend for the inclusion of data collection, data analysis and research questions in papers, there is still room for improvement suggesting that the field of real-world cybersecurity research did not mature yet. The presence of a discussion in a paper appears to affect most its citation count. However, it seems that it is not uniformly accepted what a discussion should include. This paper explores this and other issues related to paper structure and provides guidance on how to improve the quality of research reports.

Key Words: research outcome, scientific paper, article, journal paper, conference pro- ceedings, assessment, real life, natural settings, context, review, cybersecurity

Category: J.4, L.4.0, L.5

1 Introduction

Over the past decade, cybersecurity research has bloomed [Wendzel et al., 2020]. Due to its wide applicability, cybersecurity is related to various other research

fields and is being studied from several different perspectives using different research designs and methods. Among the most frequently used research designs for studying real-world phenomena are case study and action research [Fujs et al., 2019]. While action research is typically applied only in certain areas of cybersecurity, such as education, case studies appear in virtually all cybersecurity research areas [Fujs et al., 2019]. Cybersecurity may be considered as an applied discipline aiming to solve pertinent cybersecurity issues in the real-world [Bobowska et al., 2018; Tomaˇžiˇč and Bernik, 2019]. Action research and case study research designs appear to fit cybersecurity well as they focus on studying contemporary phenomena in its natural settings.

Case studies and action research have been criticized in the past [Teegavarapu and Summers, 2008; Yin, 2014; Creswell, 2014; Idowu, 2016]. Both the quality of the conducted research [Gibbert et al., 2008] and the noticeably varying quality of papers reporting on such research [Fujs et al., 2019] may be contributing to such criticism. Several approaches for evaluating the quality of research exist (e.g., [Gibbert et al., 2008; Deeks et al., 2003]) however they are time-consuming and may be relatively challenging to carry out from the outside (i.e., based on research outputs). The quality of research reports, such as scientific papers, can be essentially evaluated in two ways. The first is to assess the report impact, such as citation count and benchmarking, scholarly commentary, mass media attention, etc. [Salimi, 2017]. The second way is to evaluate the research reports themselves [Suydam, 1968]. Existing approaches for evaluation of research reports appear to be fairly time-consuming (e.g., credibility measures [Brantlinger et al., 2005]) and/or adapted for specific research designs, such as experiments [Suydam, 1968; Deeks et al., 2003]. In this paper, we focus on the evaluation of case study and action research reports to fill in an apparent lack of such evaluation approaches.

The aim of this paper is to provide a lightweight approach for evaluating case study and action research reports and apply it for evaluating real-world research in cybersecurity. To achieve this, we pose the following research questions:

RQ1: How to efficiently evaluate case study and action research reports?

RQ2: What are the trends in real-world cybersecurity research?

To answer our research questions, a novel approach for evaluating case study and action research reports is first developed. The proposed approach is then tested in a systematic literature survey. The literature survey simultaneously also offers an insight into trends in real-world cybersecurity research. This paper may benefit authors by providing guidance how to write well-structured research reports as well as the whole cybersecurity community by contributing to the

maturation of real-world research in cybersecurity.

2 Theoretical background

This paper focuses on the evaluation of two research designs, namely case study and action research. It may be important to first draw a distinction between a research design and a research method. A research design (e.g., experiment, quasi-experiment, cross-sectional, case study, action research) provides a framework for collection and analysis of data [Bryman, 2016, 2004; Parry et al., 2014]. It can be considered as a philosophy of engagement in the research process and is sometimes referred to as a research strategy [Schneider, 2012; Yin, 2014]. A research method is essentially a technique for data collection (e.g., survey, interview, observation) [Bryman, 2016, 2004; Dowling et al., 2016]. Table 1 presents four common research designs and situations in which they may be appropri- ate. The following subsections present the key characteristics of both considered research designs.

Table 1: Relevant situations for different research designs.

Research design	Type of research question	Control over events	Natural settings	Researcher influences events
Experiment	How/Why	Yes Yes/No	No	No
Quasi-experiment	How/Why	No No	Yes/No	Yes/No
Case study	How/Why		Yes	No
Action research	How/Why		Yes	Yes

2.1 Case study

Case study is a research design that is widespread across a variety of scientific fields. Although case studies often favor qualitative research methods, quantitative research methods are also frequently used [Bryman, 2016]. It may be the most appropriate research design for answering research questions with prefixes why and how, when research is focused on contemporary events, and when the circumstances do not allow for a direct, systematic and precise control over events [Yin, 2014; Eisenhardt and Graebner, 2007]. A case study is an empiri- cal inquiry that investigates a contemporary phenomenon in depth by focusing on the dynamics of the case within its real-life context and can be used even if the boundaries between the phenomenon and its context are blurred [Yin, 2014; Teegavarapu and Summers, 2008; Roth, 1999; Corcoran et al., 2004]. No- tably, case studies can be used both as a tool for testing (deductive) or building

theories (inductive) [Eisenhardt, 1989; Siggelkow, 2007; Eisenhardt and Graebner, 2007; Bryman, 2016]. Empirical scientific research should be reliable and relevant [Gibbert et al., 2008]. To establish the appropriate rigor and quality of a research design (i.e., construct validity, internal validity, external validity and reliability), four validity tests should be considered [Kidder and Judd, 1986; Yin, 2014]. Recommendations on how to achieve adequate levels of validity and reliability of case study research are presented in Table 2.

Table 2: Measures to enhance case study validity and reliability.

Validitytest	Enhancingmeasure
Construct validity	Clear chain of evidence [Yin, 2014]
Internalvalidity	Data collection and data sources triangulation [Denzin and Lincoln, 1994]
	Clearresearchframework[Eisenhardt,1989]
	Observed pattern matching with predicted ones or existing ones in literature [Eisenhardt, 1989]
	Theory triangulation [Yin, 2014]
Externalvalidity	Cross-caseanalysis[Eisenhardt,1989]
	Embedded case studies [Yin, 2014]
Reliability	Rationale for the case study selection and detailed case study context [Shadish et al., 2001]
	Detaileddescriptionofthecasestudyprotocol and study database [Yin, 2014]

A primary distinction within case study designs is between single and multiple case studies. Even though multiple case study designs are generally more compelling and considered as more robust than single case studies [Herriott and Firestone, 1983], the former may provide valuable insights, especially when studying new and emerging phenomena. There are five key rationales that justify the use of single case studies: critical case, extreme or unique case, revelatory case, representative or typical case, and longitudinal case [Yin, 2014; Bryman, 2016]. A critical case challenges an established theory in a similar way that a single experiment can (e.g., a feather free-falling in vacuum challenges the theory that the speed of free-fall depends on an object's weight). An extreme or unique case relates to a case that is very rare or unusual (e.g., the NotPetya cyberattack, 2016 US Elections). A revelatory case is based on an opportunity to study a phenomenon that is otherwise inaccessible to the scientific community (e.g., the development of Stuxnet). A representative or typical case captures the circumstances and conditions of an everyday or commonplace situation (e.g., re-

sponse to a ransomware cyberattack in a SME). A longitudinal case investigates a phenomenon over time (e.g., resilience of an organization to phishing attacks during a certain period).

2.2 Actionresearch

Action research honors the collaboration between theory and practice due to researchers and practitioners attempting to solve real-life problems together [Avison and Wood-Harper, 1986]. Unlike other research designs, it is purposely designed in a way that researchers have an influence over events and settings by providing ideas and implementing them in practice [Brantlinger et al., 2005]. Several variants of action research exist, such as participatory [Santos-Olmo et al., 2016; Schneider, 2012], collaborative [Brantlinger et al., 2005], insider [Coghian, 2001], and many others.

All variants have some common characteristics: orientation towards action and change; focus on a specific problem; a systematic and iterative process; and collaboration among all participants (i.e., researchers, practitioners and research subjects) [Santos-Olmo et al., 2016; Fletcher and Marchildon, 2014]. Action research promotes the collaboration between all participants within a commonly acceptable ethical framework [Avison et al., 1999]. It may be particularly valuable when developing new approaches and solutions [Rose, 2000; Avison and Wood-Harper, 1986].

3 A structure-based approach for evaluating case study and action research reports

The proposed structure-based approach for evaluating case study and action research reports (SAE-CSAR) draws on a typical paper structure. Scientific papers often follow the IMRaD (i.e., Introduction, Methods, Results and Discussion) structure as it helps both the authors to write relevant content to predefined parts of a paper, and the readers to seamlessly find the part of a paper that they are interested in. The introduction provides an overview of the study context and introduces the research objectives. Detailed descriptions of the research settings, data collection and analysis are provided in the methods section. The results are dedicated to presenting the outcome of data analysis and discussion to highlighting their implications and the contributions of the study. A discussion frequently evaluates the limitations of the conducted study that may provide directions for future research. For similar reasons as scientific papers follow a well-known predefined structure, several top-ranked journals (e.g., Journal of the American Medical Informatics Association (JAMIA), Journal of Knowledge Management, Journal of Enterprise Information Management) opted for structured abstracts. A structured abstract typically includes purpose / objective, research design /

methods, results / findings, discussion / implications, limitations, and originality / value.

Papers reporting on specific research designs, such as case study and action

Table 3: Evaluation criteria.

Criteria	Description	Justification
Research question	Research questions, hypotheses, a research model or other clearly stated research objectives (e.g., in a subsection).	A clear research question may help the readers in the initial stages of reading the paper to understand its (potential) contribution and may therefore influence their decision whether to continue reading until the end.
Case description	A description of the unit of analysis and its context. A well-written case description may add credibility to the paper as well as help the readers to grasp the scope of the study.	Introduces the reader to the context of the research study.
Data collection	Reported data sources and procedures (i.e., which data was collected and how).	Adds to the credibility of the report by detailing which data was collected and how.
Data analysis	Described data analysis procedures and tools used.	Facilitates the understanding of how the data was analyzed.
Ethical considerations	Notes on issues related to research ethics (e.g., ethical board approval). Relevant only for studies involving human beings (e.g., interviews, surveys) and may therefore not be relevant for all real-world research.	It adds to the credibility of the study. However, it may also set some ethical flexibility as studies need to be well-planned in advance (e.g., not to wander out of the scope of an ethical board approval).
Results	Presented data analysis results.	Well-presented results may provide a credible base for the discussion.
Discussion	A discussion beyond a description or simple interpretation of the results (e.g., theoretical and practical implications of results).	Highlighted key contributions and implications of the study may help the readers to determine the value of the study.
Limitations	Considered limitations of the study.	Most likely, all research has limitations. Highlighting the most important ones may help the readers to better understand the scope of the study contribution and potential directions for future work.

research, may further adapt the standard paper structure. First, research objectives are commonly defined with research questions in case study and action research as opposed to hypotheses and research models that are more common in other research designs (e.g., experimental, cross-sectional design). Next, a case description in case study and action research is essentially the description of research settings. Finally, case studies and action research regularly involve interaction with people (e.g., during data collection with interviews) or otherwise influence them (e.g., by conducting a phishing campaign on an organization [Mihelič et al., 2019]). Therefore, ethical considerations may be relevant, especially for real-world research.

Building on the above considerations, we developed a novel approach for evaluating case study and action research reports. The proposed approach is based on eight key paper parts: research question, case description, data collection, data analysis, ethical considerations, results, discussion and limitations. The evaluation criteria with a brief description and justification are presented in Table 3.

Evaluation of case study and action research reports with the proposed approach is quite straight-forward. Criteria scores are determined based on whether the paper includes an evaluation criteria part or not. For example, if the paper reports the results, the score for the *results* criteria is 1 otherwise it is 0.

4 Methods

A systematic literature survey of papers reporting on real-world research in cybersecurity was conducted to (1) evaluate the feasibility of the proposed approach for evaluating case study and action research reports and (2) to gain an insight into the trends in real-world cybersecurity research.

4.1 Data collection

The survey covered conference and journal papers published from 2015 to 2019 spanning over the period of 5 years. Figure 1 presents the literature survey
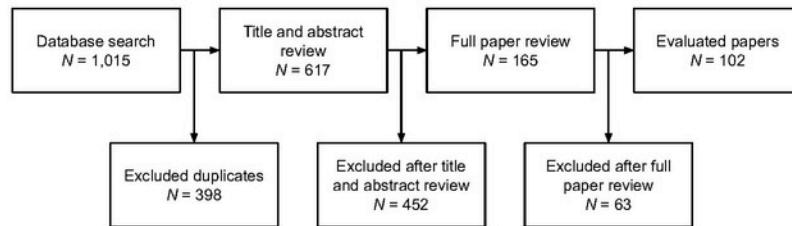

Figure 1: Literature survey process.

Table 4: Search queries, filters and the number of returned bibliographic records.

Database	Queries / Filters / Returned records
ACM DL	51 records
Period	2015 – 2019
Type	Research article
Query	Title:(("cyber security" OR "cybersecurity") AND ("case study" OR "action research")) OR Abstract:(("cyber security" OR "cybersecurity") AND ("case study" OR "action research")) OR Keyword:(("cyber security" OR "cybersecurity") AND ("case study" OR "action research")) OR Title:("cyber security" OR "cybersecurity") AND Abstract:("case study" OR "action research") OR Title:("case study" OR "action research") AND Abstract:("cyber security" OR "cybersecurity") OR Title:("cyber security" OR "cybersecurity") AND Keyword:("case study" OR "action research") OR Title:("case study" OR "action research") AND Keyword:("cyber security" OR "cybersecurity") OR Abstract:("cyber security" OR "cybersecurity") AND Keyword:("case study" OR "action research") OR Abstract:("case study" OR "action research") AND Keyword:("cyber security" OR "cybersecurity")
IEEE Xplore	201 records
Period	2015 – 2019
Type	Conferences, Journals
Query	(("All Metadata":"cyber security" OR "cybersecurity") AND ("All Metadata":"case study" OR "action research"))
Scopus	505 records
Period	2015 – 2019
Type	Article or Conference paper
Query	TITLE-ABS-KEY(("cyber security" OR "cybersecurity") AND ("case study" OR "action research"))
Web of Science	258 records
Indexes	SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC
Period	2015 – 2019
Type	Proceedings paper, Article
Query	TOPIC:(("cyber security" OR "cybersecurity") AND ("case study" OR "action research"))

process with the number of included papers in each step.

First, four bibliographic databases (i.e., ACM DL, IEEE Xplore, Scopus and Web of Science) were queried on 4 January 2020. The search queries included a combination of keywords cyber security, cybersecurity, case study and action research. The queries were used to search the bibliographic records (i.e., title, abstract and keywords) and differ between the selected bibliographic databases. Table 4 presents the search queries, applied filters and the number of returned

bibliographic records for each individual bibliographic database. A total of 1,015 bibliographic records were returned by querying the bibliographic databases. Second, duplicate records were removed resulting in 617 unique bibliographic records. Third, two researchers separately examined the title and abstract of each record to determine whether to include it in the survey according to inclusion and exclusion criteria presented in Table 5.

Table 5: Inclusion and exclusion criteria

Inclusioncriteria	Exclusioncriteria
Real-worldcasestudy	Notdirectlyrelatedtocybersecurity
Real-worldactionresearch	Non-real-worldcasestudy
Journalorconferencepaper	Notcasestudyoractionresearch
WritteninEnglish	Editorial,bookchapter,posterorinvitedtalk
Published between 2015 and 2019	Work in progress
	Full text not available to researchers

Case study is an ambiguous term and has various meanings in research. First, case studies may refer to purely theoretical (e.g., mathematical) research. For example, Standley and Boucheron [2018] use GPS as a case study to present a mathematical formulation of the contribution of space-based unidirectional systems to the resilience of strategic cyber networks. Second, real-world data or subjects can be used in artificial settings. For example, Genç et al. [2019] present a case study of a real-world subject (i.e., ransomware). However, the subject is observed in an artificially created environment (i.e., a research testbed) which suggests an experimental or quasi-experimental research design. It is similar with research using real-world data in artificially created settings. Third, case study is sometimes used as a synonym for a use case. A use case considers the application to a specific domain. However, there is no specific case being studied per se and research often describes a theoretical instead of a real-world application. Fourth, case studies are occasionally wrongly reported when a cross-sectional design has been employed. A cross-sectional design is often also called a survey design which is its most common form. It entails the collection of quantitative or quantifiable data on (much) more than one case at a single point in time [Bryman, 2016]. For example, Harrell et al. [2018] reported case studies of higher education institutions. Data collection focused on 272 institutions at a certain point in time and provided an insight into the overall situation. However, little to nothing is known about any of the 272 individual cases (i.e., higher education institutions) and therefore cannot be considered having a case study research design. Other examples of declared case studies with similar issues include surveys in a specific context (e.g., public sector in Zambia [Chinyemba and Phiri,

2018]) and live classification of real-world data (e.g., Twitter profiles [Tundis et al., 2018; Tundis and Muhlhauser, 2017]).

Inconsistencies between the researchers' assessments were identified and solved with consensus between the two researchers. In cases when it was not clear from the title and abstract whether to keep a record for further anal- ysis, the full paper was briefly examined. If the researchers determined that a paper employed a different research design (e.g., action research) instead of the declared research design (e.g., case study) due to obvious reasons (e.g., the influence of the researchers on the case), the paper was reclassified accordingly. After applying the inclusion and exclusion criteria, 165 records were retained for full paper analysis.

Fourth, all retained papers were downloaded and evaluated according to the evaluation criteria of the proposed approach. Two researchers first individually evaluated papers according to the evaluation criteria. Inconsistencies between scores were then identified and solved with consensus between the researchers. Papers were additionally excluded according to the inclusion and exclusion criteria in this step if the researchers determined this was necessary after a detailed examination of the paper. After the quality evaluation, N = 102 records (Ncs = 91, Nar = 11) were finally retained in the literature survey. For these records, citation counts in all bibliographic databases were obtained. Citations in ACM DL were obtained on 9 January 2020 while citations in other databases were obtained on 4 January 2020. The paper citation count was determined by the highest number of citations in any of the bibliographic databases.

4.2 Dataanalysis

First, keywords were analyzed with word clouds to gain an insight into the trends in real-world cybersecurity research over the studied period. Word clouds are a form of text visualization that enriches the primary data in a meaningful way and enables new ways to interpret the data [Jayashankar and Sridaran, 2017]. Keyword clouds were created with a free online application (www.wordclouds.com). Second, the results of the evaluation of papers were visualized using column charts which were interpreted to determine potential associations with other collected bibliometric data. Charts were created with Google Sheets.

5   Results

This section first presents the trends in real-world cybersecurity research for the studied period. Next, paper evaluation results are presented and potential associations to other bibliometric data are identified.

5.1 Real-world research in cybersecurity

Figure 2 presents keyword clouds for each year in the studied period. The size of a keyword is proportional to frequency of its occurrence and the sequence of keyword clouds visualizes trends in the usage of keywords in the studied area. Keywords "cyber", "security" and "cybersecurity" are consistently among the most frequent keywords in the studied papers. The keyword "cybersecurity" gradually gained popularity and became one of the most frequently used keywords only in 2017.

(a) 2015   (b) 2016   (c) 2017
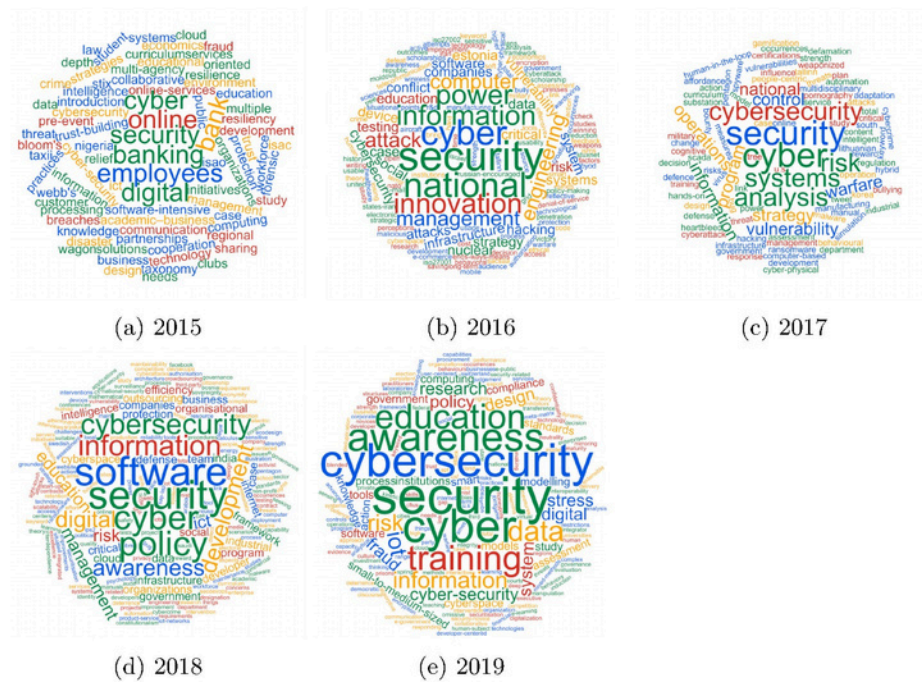
(d) 2018   (e) 2019

Figure 2: Keyword clouds for 2015 – 2019.

The other keywords point to the trends in published topics. In 2015, the most common keywords were "banking", "online", "employees", "digital" and "bank". This suggests a focus on the cybersecurity of organizations, especially financial institutions. In 2016, the keywords "innovation", "national", "information" and "power" point to the emergence of cybersecurity on the national level. More precisely, critical infrastructure protection and case studies of national cybersecurity powers appear to dominate other studied topics. The trend turned back to organizations but without the focus on financial institutions in 2017

with the most frequent keywords being "risk", "analysis" and "systems". The topics seems to have switched towards software-related cybersecurity solutions in 2018 with the keyword "software" visibly standing out. Other frequent keywords include "information" and "policy" suggesting the continuation of focusing on organizations. In 2019, a major shift in topics occurred with most common keywords being "awareness", "education" and "training". This may be a response to the persistence of one of the weakest links in cybersecurity, namely people.

5.2 Evaluation of case study and action research reports

Papers were evaluated according to evaluation criteria of the proposed approach in two steps. First, researchers individually scored the paper. Second, inconsistencies were identified and resolved through consensus. Overall, the proposed approach appears to be relatively efficient since only 55 out of 516 scores (6.7 percent) were inconsistent in the first step. Figure 3 shows the breakdown of these inconsistencies according to the evaluation criteria. The most challenging task in evaluating papers appears to be determining whether a paper includes a discussion or not as 18.6 percent of papers were scored inconsistently.
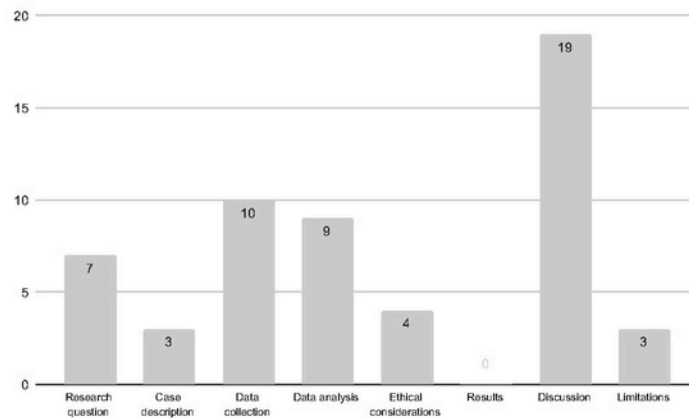

Figure 3: Inconsistencies between researchers' evaluation criteria scores.

The overall summary of paper evaluation is presented in Figure 4. The data suggests that there is a significant body of poorly rated conference papers with only 2 or 3 evaluation criteria parts. Most of papers with 2 or 3 evaluation criteria parts report *results* and *case description*. The third reported part is either *data collection* (a third of papers), *discussion* (a fifth of papers) or *research question*

(a tenth of papers). *Data analysis, ethical considerations* and *limitations* are reported in a negligible number of these papers.
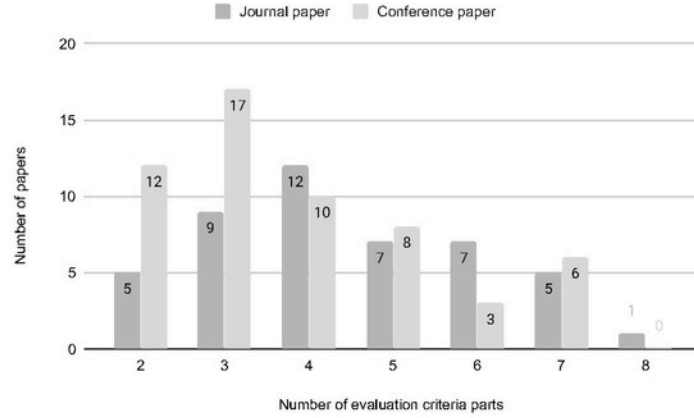

Figure 4: Histogram of evaluation criteria parts in journal and conference papers.

Figure 5 shows the share of evaluated papers according to individual evaluation criteria parts. Less than half of the evaluated papers include a *research question* or otherwise formally defined research objectives. In some papers, the research question is emphasized and easy to find. However, in other papers it is integrated into text. At least a brief *case description* is provided in the vast majority of papers. The descriptions vary considerably in how detailed they are regarding both the unit of analysis and its context and can take from a few sentences to whole subsections. *Data collection* is reported in approximately two thirds of papers. A minimum data collection presentation enables at least the identification of data sources used. *Data analysis* is included in less than half of publications and is occasionally integrated with the presentation of results. Nevertheless, a significant share of papers only reports the results without giving any insight into how they were reached. All papers reported at least some *results* without exception. There are, however, considerable differences in the amount and quality (e.g., relevance, presentation, completeness) of the presented results. A negligible number of papers report on the *ethical considerations*. In some cases, ethical considerations are not relevant as research does not involve people. However, a considerable share of real-world research involves collecting data from people (e.g., interviews, surveys) implying that ethics should be considered more often. *Discussion* is covered in a slight majority of journal papers but only in a third of conference papers. Discussion is often supposedly integrated with the

results, however, often lacks a discussion beyond a simple interpretation of the results and does not provide any implications nor contributions of the paper. Sometimes, a proper discussion is provided in the conclusion. *Limitations* are the second least reported part of a paper despite their importance for gaining a more objective grasp on the contribution of a study. The key differences between journal and conference papers are in the share of papers with discussion (23.0 percent), research question (15.3 percent), limitations (14.8 percent) and data analysis (12.1 percent). These seem to be the paper parts that are more often required by journals than conferences.
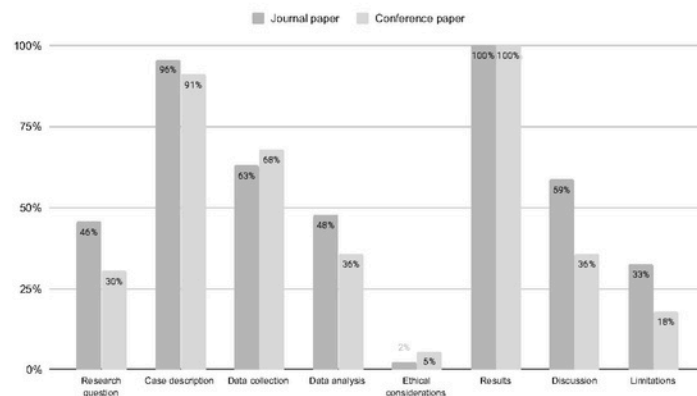

Figure 5: Share of journal and conference papers satisfying the evaluation criteria.

Figure 6 presents the share of papers reporting individual evaluation criteria parts for the studied period. There is a noticeable upward trend for reporting data collection in both journal and conference papers. Research question and data analysis are also included in papers more frequently in recent years. Ethical considerations are reported in conference papers more often since 2018 (13 and 5 percent in 2018 and 2019, respectively) while journal papers only reported them in 2016 (14 percent). There is a slight downward trend for including case description and discussion in papers. Despite an upward trend for reporting discussion in conference papers, the noticeable downward trend in journal papers overshadows it. At its peak in 2016, 86 percent of journal papers included a discussion. The share fell sharply each year to 36 percent in 2019 which may be partially attributed to the doubling of the number of journal papers in the same period (e.g., more publication venues with varying research rigor requirements). These results indicate that there is still room for improvement and that real-

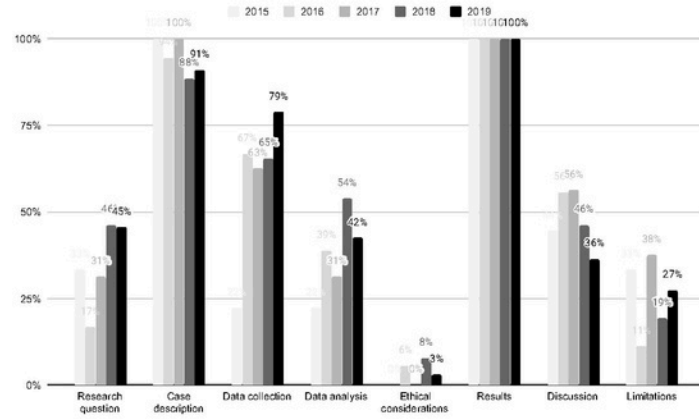world research in cybersecurity did not reach its maturing point yet.

Figure 6: Share of journal and conference papers satisfying the evaluation criteria for 2015 − 2019.

Figure 7 compares average citations for papers with or without individual evaluation criteria to uncover which parts of the paper impact most the citations. The most important part of the paper appears to be the discussion. Average citations of papers with discussion are more than 3 times higher
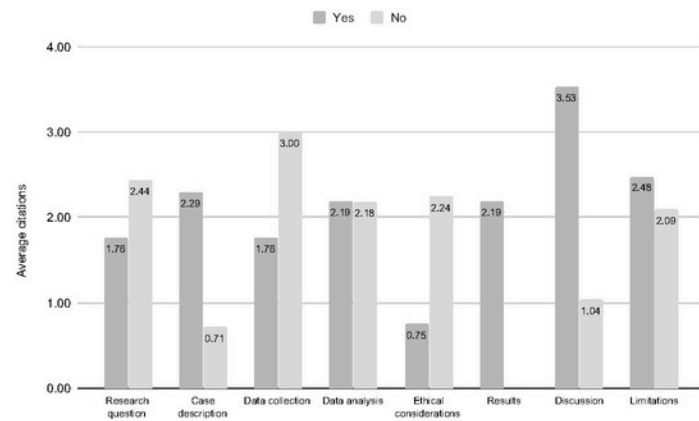

Figure 7: Average citations per evaluation criteria.

than average citations for papers without. It is similar with case descriptions. However, this comparison may be unreliable due to a low number of papers without case description. For similar reasons, we cannot say much about the importance of results and ethical considerations. There are only slight differences in average citations with and without data analysis and limitations suggesting that researchers tend not to consider them when deciding to cite a paper. It is interesting to note that papers that do not report the research question or data collection have higher average citations than those that do. If we assume that researchers do not search for papers that are missing these parts, this may indicate the noise level in our sample. If this is the case, only the presence of the discussion may affect average citations.

6 Discussion

This paper makes several theoretical and practical contributions described in the next subsections. The paper concludes with a discussion on limitations of the study and potential avenues for future work.

6.1   Theoretical implications

First, this paper develops a novel approach for evaluating case study and action research and applies it in a literature survey of real-world research in cyber-security to determine its usefulness. Although the proposed approach does not evaluate the quality of the papers in detail, it appears to offer a reasonable alternative to existing comprehensive evaluation approaches that require more time for the evaluation of a paper. Therefore, the key merit of the proposed approach is that it enables gaining a thought-provoking insight into the quality of papers in a studied research area in an efficient way.

Second, the approach seems to be appropriate for evaluation of case study and action research reports in a research area that has not matured yet (i.e., a significant share of papers is missing some key evaluation criteria parts). In our study, results and case description were included in almost all studied papers. This suggests that some adjustments to the proposed evaluation model may be needed, especially as the research area matures and papers include more evaluation criteria parts. For example, an ordinal (e.g., levels of quality with descriptions) or interval scale (e.g., from 1 to 5) may be used to evaluate the criteria instead of the binary one. Alternatively, some evaluation criteria may break down into more detailed sub-criteria. For example, the case description could be further divided to explicitly defining the unit of analysis, description of the unit, description of the context, etc. although this would decrease the efficiency of the approach.

Third, 63 out of 165 fully reviewed papers were excluded (38.2 percent). This suggests that abstracts may not be informative enough. One of the reasons is that real-world research is often declared in the abstract despite that not being the case (e.g., real-world data being used in artificial settings). Since this study does not focus on evaluation of abstracts, it would be beneficial if future studies would address this issue.

Fourth, the trends in the usage of keywords points to a wide array of different topics from technological topics to management, state cybersecurity and human-related topics, such as awareness, education and training. This firmly establishes the inter-disciplinary nature of cybersecurity. Although most cybersecurity research may be on technological topics, these studies are typically conducted in artificial settings. Therefore, there appears to be a research gap and opportu- nity for future research, for example, in real-life applications of technological cybersecurity solutions.

6.2   Practical implications

First, this paper provides a relatively light-weight approach for evaluating case study and action research reports. Although it has been applied in the context of real-world research in cybersecurity, the proposed approach is universal and can be applied to any context. A drawback of the proposed approach might be that it may not be appropriate when surveying a mature research area. To overcome this deficiency, the approach may be extended by using more detailed scales for scoring or expanding the evaluation criteria.

Second, although the proposed approach is primarily intended for evaluation of research reports, this paper may also serve as s guidebook for authors and reviewers. By using the proposed approach as a checklist, the authors may be able to better structure their papers and improve their readability. Reviewers may contribute to better quality of publications and maturation of the research field by requesting the authors to clearly structure their papers.

Third, a fairly straight forward practical implication is that authors should put more attention to the discussion. A high number of inconsistencies between individual scores of researchers, and the varying location of implications and contribution in papers may indicate that a "discussion" has different meanings for different authors. Putting the study implications and paper contribution in a standard location, namely the discussion, may help the readers to quickly find the added value of the paper and its justification. This may be important due to the potential effect of the discussion on paper citations that we detected in our study.

7 Limitations and future work

Like most research, this paper has some limitations that the readers should consider. First, only two researchers evaluated the papers independently. One researcher has a background in computer science and the other one in social sciences. Involving more researchers with diverse backgrounds would help to further validate the proposed approach.

Second, this study shows that this approach can provide a thought-provoking insight into the studied field by considering 102 papers from a 5-year period. Including less papers may affect the usefulness of the proposed approach for providing meaningful insight into the studied field. Similarly, the usefulness of the proposed approach may be affected if the number of papers per year or type is too low. Since the minimum of papers required for this approach to be useful is hard to determine on this study alone, future studies employing the proposed approach may help in establishing the prerequisites for using the proposed approach.

Third, the proposed approach has been applied to a research area that has not fully matured yet. Insights from applications to a more mature research area would be highly beneficial and would provide some hints on whether the proposed approach can be useful in its original form also in those settings.

Fourth, this paper explored the association between different parts of the paper and paper citations. However, the paper abstract and its quality may also have significant effects on citations as researchers typically read the abstract first. A systematic evaluation of paper abstracts would be valuable in determining these effects. An analysis of the association between the quality of an abstract and paper quality would also be interesting.

Fifth, the number of evaluated papers was not large enough to conduct more rigorous statistical tests, such as logistic regression. For example, to determine which evaluation criteria affects the number of citations (e.g., top 20 percent best cited papers compared to the rest), at least approximately 500 papers would need to be evaluated.