ORIGINAL ARTICLE

Performance in Command and Control: Results From a Scoping Review

Patrick Drews[1] (ORCID) · **Frank Fiedrich[1]**

Abstract

Command and control are critical components of emergency management when disaster strikes. Command posts act as a support system within the command chain for leaders, and the performance of these teams is essential for the successful operation. This paper presents the findings of a scoping review on the performance of command posts, identifying six performance factors: scenario, resilience, situational awareness, decision making, team structures and teamwork, and operational execution. Furthermore, this paper presents objective indicators for performance measurement in three dimensions: process, effectiveness, and efficiency. Additionally, it provides insights and methodologies for evaluating command post exercises.

Keywords Command and control · Command post · Exercise · Performance

1 Introduction

Disasters, major fires, and the defuse of unexploded artillery left over from World War II in urban areas are just a few examples of situations in which first responders establish coordinating committees. In Germany, police, fire fighters, rescue services, customs and comparable organizations are known as authorities and organizations with security tasks (Behörden und Organisationen mit Sicherheitsaufgaben, BOS). These coordinating committees are known in technical jargon as command and control (C2) teams or command posts—the German term is "Stab". Their task is to collect information, evaluate that information, and prepare decisions to support the incident commander in ambiguous or dynamic incidents.

✉ Patrick Drews
  patrick.drews@uni-wuppertal.de

  Frank Fiedrich
  frank.fiedrich@uni-wuppertal.de

1   Institute for Public Safety and Emergency Management, University of Wuppertal, Gaussstr. 20, 42119 Wuppertal, Germany

🌱 Springer

Command and control teams are particularly active when many responders from different organizations are on scene or when many incident sites are occurring, for example, as a result of severe weather events. The command posts consist of experienced managers who support the incident commander. The German emergency management doctrine 'Dienstvorschrift 100' (DV100) (Ausschuss Feuerwehrangelegenheiten, Katastrophenschutz und zivile Verteidigung (AFKzV) 1999) distinguishes between the operational level ("operativ-taktisch") and the administrative level ("politisch-administrativ"), with both holding equal power under the command of the administrative head ("Hauptverwaltungsbeamter"). Operational command posts are primarily staffed by fire department personnel.

1.1 Rationale

The frequency of deployment of command posts within the German emergency management system is not completely determined (Rönnfeldt 2012) and depends on the frequency of incidents and the cause for the deployment of the emergency management authority. However, the statistics for the deployments are missing. According to Lamers (2016), there is an annual frequency of one exercise and approximately five training sessions for command posts at the district level in North Rhine-Westphalia (one of the 16 German states). Following this estimate, it can be concluded that the deployment of a command post at the district level is highly irregular and ad hoc, occurring only once every few years. The irregular deployment and limited daily involvement require consistent training in the necessary skills. The development of competence in crisis management is considered a knowledge gap, as managers lack the necessary skills in case of an emergency (Cerqueira et al. 2017). In the field of civil protection and emergency management, practical experience is of great importance (Hufschmidt and Dikau 2013). To maintain the required level of expertise, the team members receive initial and advanced training courses and participate in practical exercises. Simulation events and the subjecting of individuals to progressive stress is a suitable technique for training individuals for leadership selection (Cerqueira et al. 2017). Furthermore, exercises can provide valuable operational experience for staff (Mitschke and Karutz 2017).

Exercises are defined as "*the process to train for, assess, practise, and improve performance in an organization*" (DIN EN ISO 22300 2021). This publication will focus on the assessment aspect of exercises.

Command and control is defined as "*The exercise of authority and direction by a commander over assigned, allocated and attached forces in the accomplishment of a mission*" (Bernier et al. 2012). In a scoping review related to disaster exercise evaluations, Beerens and Tehler (2016) identified them as "*lessons learned areare primarily derived from observations or collected after debriefings*". They concluded that the field lacks a comprehensive knowledge base. In a follow-up study Beerens (2019) analyzed evaluation documents from the Netherlands, concluding "*that there is a need for empirical research that supports evaluation practice.*" Both studies had a broader view on the field of exercise evaluation, which leads to the aim of this study.

1.2 Objectives

The aim of this study is to assist in evaluation by providing a deeper understanding of what is known in the scientific community about performance in command post exercises. This research is part of a dissertation at the University of Wuppertal, with the objective of developing key performance indicators to evaluate command post exercises.

2 Methods

The presented study was conducted from April 2021 to July 2023 following a scoping approach.

> "A scoping review or scoping study is a form of knowledge synthesis that addresses an exploratory research question aimed at mapping key concepts, types of evidence, and gaps in research related to a defined area or field by systematically searching, selecting and synthesizing existing knowledge" (Colquhoun et al. 2014).

For this research, the scoping approach is appropriate because it "addresses broader topics where many different study designs might be applicable"(Arksey and O'Malley 2005).

This research follows the "6-stage framework" of scoping reviews (Arksey and O'Malley 2005; Levac et al. 2010) supported by the search protocol by RefHunter (Nordhausen and Hirt 2020). The first four stages are described in the following subsections, while stage 5 ("Collating, summarizing, and reporting the results") and stage 6 ("Consultation") will be represented by the results (Sect. 3) and discussion (Sect. 4) sections of this paper.

2.1 Stage 1: Identifying the Research Question

In scoping studies, it is preferable to pose a broad research question to encompass various scientific publications. However, there is a risk of confusion. Levac et al. (2010) recommend taking into account the intended outcome, the rationale, and the purpose for improved precision. For this reason, the broad research question was supported by additional and more specific questions. The broad research question is: What is known in the scientific community about the performance of command posts? The supporting questions are as follows: (a) What defines performance in command and control? (b) What are the criteria to measure command and control performance? Additionally, (c) How are performance evaluations conducted during exercises? Before the systematic search, which is represented by stage 2, several preliminary searches were conducted to gain an understanding of the results yielded by the database query.

2.2 Stage 2: Identifying Relevant Studies

To identify the studies relevant to the research question, the identification process was divided into two parts. Firstly, a selection of relevant databases ('where to search') and secondly, the development of search terms to query the identified databases ('how to search').

2.2.1 Databases Selection

For the selection of relevant databases, a search was conducted within the database information system DBIS by the University Library of Regensburg (Universitätsbibliothek Regensburg 2023). This search led to the selection of five databases: Scopus, Web of Science, and Thomson Reuters were chosen because they are multidisciplinary databases that cover numerous research fields and topics, and are owned by various publishers. Moreover, the PubMed database by the United States National Center for Biotechnology Information was selected because it encompasses a broad range of publications from the medical sciences. Finally, CORDIS and TibKat databases were chosen for the following reasons: The European Commission and the German Federal Ministry of Education and Research are funding projects regarding the coordination of first responders, the focus of this study. These projects are required to disseminate their research findings through non-traditional academic means (e.g. deliverables or research reports). The CORDIS database contains the research results from projects funded by the EU. Similarly, the library at the Technical University of Hannover houses the reports for projects funded in Germany by the Federal Ministry of Education and Research (BMBF).

2.2.2 Search Query Development

The database queries utilized a Boolean logic approach, combining multiple keywords with bool operators. Four search components were identified based on the research questions: (1) emergency response and military forces, (2) performance characteristics and measurement methods, (3) measurability of command and control / C2 teamwork, and (4) exercises and operations. Search terms were derived from these components. Because the research focus of the dissertation is the German 'Stab,' we chose German and English terms that were translated in both directions. No publication time restriction was applied. The search terms are presented in Table 1.

Initial searches carried out in stage 1 revealed various studies that examined the impact of exercises on medical conditions (such as heart diagnostics), sports performance, and skill-specific training (such as a neurosurgery procedure). However, studies related to therapy or physical work were excluded through the use of stop terms because they did not cover the research questions presented in Sect. 2.1. Finally, the search terms were converted into queries for each designated database. The queries are displayed in Table 2.

Scopus and PubMed were searched on March 11, 2021, while the remaining databases were searched on May 29, 2021. These dates were selected based on the

Table 1 Search components

Search components	Search terms (German)	Search terms (English)
Search component 1: emergency response and military forces	EINSATZKRÄFTE MITARBEITER*INNEN GEFAHRENABWEHR EINSATZ	EMERGENCY PERSONNEL EMERGENCY FORCES EMERGENCY RESPONSE DISASTER RESPONSE EMERGENCY SERVICE
	MILITÄR	MILITARY ARMED FORCES
	KATASTROPHENSCHUTZ ZIVILSCHUTZ	CIVIL PROTECTION DISASTER PROTECTION
	STAB FÜHRUNGSGRUPPE FÜHRUNGSGREMIUM FÜHRUNGSSTAB	STAFF COMMANDING UNIT COMMAND POST TASK FORCE
Search component 2: performance characteristics and measurement methods	LEISTUNG	PERFORMANCE CAPACITY OUTPUT
	MESSUNG EVALUATION BEWERTUNG AUSWERTUNG	MEASUREMENT MEASURING RATING ASSESSMENT EVALUATION ANALYSIS VALUATION REVIEW
	MESSVERFAHREN MESSMETHODE	MEASUREMENT METHOD MEASUREMENT PROCEDURE MEASUREMENT PROCESS
	KEY PERFORMANCE INDICATORS QUALITÄTSINDIKATOREN	KEY PERFORMANCE INDICATORS QUALITY INDICATORS
Search component 3: measurability of command and control/C2 teamwork	FÜHRUNGSARBEIT	LEADERSHIP
	KOORDINATION	COORDINATION MEASURABILITY
	MESSBARKEIT	COMMAND & CONTROL
Search component 4: exercises and operations	ÜBUNG ÜBUNGEN TRAINING	EXERCISE EXERCISES DRILL SIMULATION TRAINING
	EINSATZ	MISSION DEPLOYEMENT

Table 2 Search queries

Database	Query
Scopus	TITLE-ABS-KEY (EINSATZKRÄFTE OR MITARBEITER* OR GEFAHREN-ABWEHR OR MILITÄR OR KATASTROPHENSCHUTZ OR ZIVILSCHUTZ OR STAB OR FÜHRUNGSGRUPPE OR FÜHRUNGSGREMIUM OR FÜHRUNGSSTAB OR "EMERGENCY PERSONNEL" OR "EMERGENCY FORCES" OR "EMERGENCY RESPONSE" OR "DISASTER RESPONSE" OR "EMERGENCY SERVICE" OR MILITARY OR "ARMED FORCES" OR "CIVIL PROTECTION" OR "DISASTER PROTECTION" OR STAFF OR "COMMANDING UNIT" OR "COMMAND POST" OR "TASK FORCE") AND TITLE-ABS-KEY (LEISTUNG OR MESSUNG OR EVALUATION OR BEWERTUNG OR AUSWERTUNG OR MESSVERFAHREN OR MESSMETHODE OR "KEY PERFORMANCE INDICATORS" ORQUALITÄTSINDIKATOREN OR PERFORMANCE OR CAPACITY OR OUTPUT OR MEASUREMENT OR MEASURING OR RATING OR ASSESSMENT OR EVALUATION OR ANALYSIS OR EVALUATION OR REVIEW OR "MEASUREMENT METHOD" OR "MEASUREMENT PROCEDURE" OR "MEASUREMENT PROCESS" OR "KEY PERFORMANCE INDICATORS" OR "QUALITY INDICATORS") AND TITLE-ABS-KEY (FÜHRUNGSARBEIT OR KOORDINATION OR MESSBARKEIT OR LEADERSHIP OR COORDINATION OR MEASURABILITY OR "COMMAND &CONTROL" OR "COMMAND AND CONTROL") AND TITLE-ABS-KEY (ÜBUNG* OR TRAINING OR EINSATZ OR EXERCISE* OR DRILL OR SIMULATION OR TRAINING OR MISSION OR DEPLOYEMENT)
PubMed	(((((("EMERGENCY PERSONNEL") OR ("EMERGENCY FORCES") OR ("EMERGENCY RESPONSE") OR ("DISASTER RESPONSE") OR ("EMERGENCY SERVICE") OR (MILITARY) OR ("ARMED FORCES") OR ("CIVIL PROTECTION") OR ("DISASTER PROTECTION") OR (STAFF) OR ("COMMAND POST") OR ("TASK FORCE")) AND ((PERFORMANCE) OR (CAPACITY) OR (OUTPUT) OR (MEASUREMENT) OR (MEASURING) OR (RATING) OR (ASSESSMENT) OR (EVALUATION) OR (ANALYSIS) OR (EVALUATION) OR (REVIEW) OR ("MEASUREMENT METHOD") OR ("MEASUREMENT PROCEDURE") OR ("MEASUREMENT PROCESS") OR ("KEY PERFORMANCE INDICATORS") OR ("QUALITY INDICATORS") OR ("HEALTH SERVICES") OR ("DIAGNOSTIC TESTS") OR (ROUTINE) OR ("RESEARCH DESIGN") OR ("QUALITY INDICATORS") OR ("HEALTH CARE")) AND ((LEADERSHIP) OR (COORDINATION) OR (MEASURABILITY) OR "(COMMAND & CONTROL") OR ("COMMAND AND CONTROL") OR ("DISASTER PLANNING")) AND ((EXERCISE*) OR (EXERCISES) OR (DRILL) OR (SIMULATION) OR (TRAINING) OR (MISSION) OR (DEPLOYMENT) OR (EXERCISE) OR ("SIMULATION TRAINING"))) NOT (THERAPY) NOT (PHYSICAL) NOT (CANCER[MESH]) AND ("TASK PERFORMANCE AND ANALYSIS"[MESH] OR "WORK PERFORMANCE"[MESH] OR "STANDARDS" [SUBHEADING] OR "QUALITY IMPROVEMENT"[MESH] OR "PROCESS ASSESSMENT, HEALTH CARE"[MESH] OR "SYSTEMS ANALYSIS"[MESH] OR "SELF-EVALUATION PROGRAMS"[MESH] OR "HIGH RELIABILITY ORGANIZATIONS"[MESH])
TibKat	(SUPPLIERPREFIX:TIBKAT) AND (DOCUMENTTYPE:R) AND (EINSATZ-KRÄFTE OR MITARBEITER* OR GEFAHRENABWEHR OR MILITÄR OR KATASTROPHENSCHUTZ OR ZIVILSCHUTZ OR STAB OR FÜHRUNGS-GRUPPE OR FÜHRUNGSGREMIUM OR FÜHRUNGSSTAB)

Table 2 (continued)

Database	Query
Web of Science	ALL=(((EMERGENCY) OR ("CIVIL PROTECTION") OR ("TASK FORCE") OR (MILITARY) OR ("COMMANDING UNIT")) AND ((PERFORMANCE) OR (MEASUREMENT) OR (OUTPUT) OR (ASSESSMENT) OR (EVALUATION) OR (ANALYSIS) OR ("KEY PERFORMANCE INDICATORS") OR (REVIEW) OR ("QUALITY INDICATORS")) AND ((LEADERSHIP) OR (COORDINA-TION) OR (MEASURABILITY)) AND ((EXERCISE*) OR (DRILL) OR (SIMULATION) OR (TRAINING) OR (MISSION) OR (DEPLOYMENT))) REFINED BY: [EXCLUDING] DOCUMENT TYPES: (MEETING ABSTRACT OR EDITORIAL MATERIAL OR BOOK REVIEW) TIMESPAN: ALL YEARS. INDEXES: SCI-EXPANDED, SSCI, A &HCI, CPCI-S, CPCI-SSH, ESCI
Cordis	/RESULT/RELATIONS/CATEGORIES/COLLECTION/CODE='DELIVERABLE' ,'PUBLICATION','EXPLOITABLE' OR /RESULT/RELATIONS/CATEGORIES/ COLLECTION/CODE='PUBSUM') AND (('EMERGENCY' OR 'CIVIL PRO-TECTION' OR 'TASK FORCE') AND ('EXERCISE' OR 'SIMULATION'))
BASE[1]	"EMERGENCY MANAGEMENT" AND PERFORMANCE AND TRAINING

[1]Additional key words yielded in 0 results

availability of the main author. Access to TibKat was temporarily unavailable and required resolution through direct communication with the provider. The search queries yielded 11,147 results, which were subsequently analyzed.

2.3 Stage 3: Study Selection

The study analyzes articles related to command post performance. The areas of interest were identified in four domains based on the research questions: (1) Command posts / teams with C2-related tasks[1], (2) exercises and missions, (3) performance characteristics, and (4) evaluation methods. The articles were included if they encompassed information on command posts in emergency management or military exercises, simulated or real-world missions, case reports of missions, performance characteristics, measurement criteria, measurement methods, observations, numerical measurement methods and methods using sensors. Table 3 presents inclusion and exclusion criteria.

The Distiller SR Suite by DistillerSR Inc. was used for the screening process. After uploading the search results to the database, an automated duplicate check was run, comparing DOI, PMID, and PID. Duplicates that were missed in the initial screening were removed during subsequent screenings, the first of which reviewed the publication titles and the second of which focused on abstracts. As a result, 9,171 records were excluded. Titles or abstracts that were missing or that led to uncertainty about their eligibility were included in the final analysis. The

---

[1] Command posts is a military term. To include non-military command and control, a more broadly defined term was used.
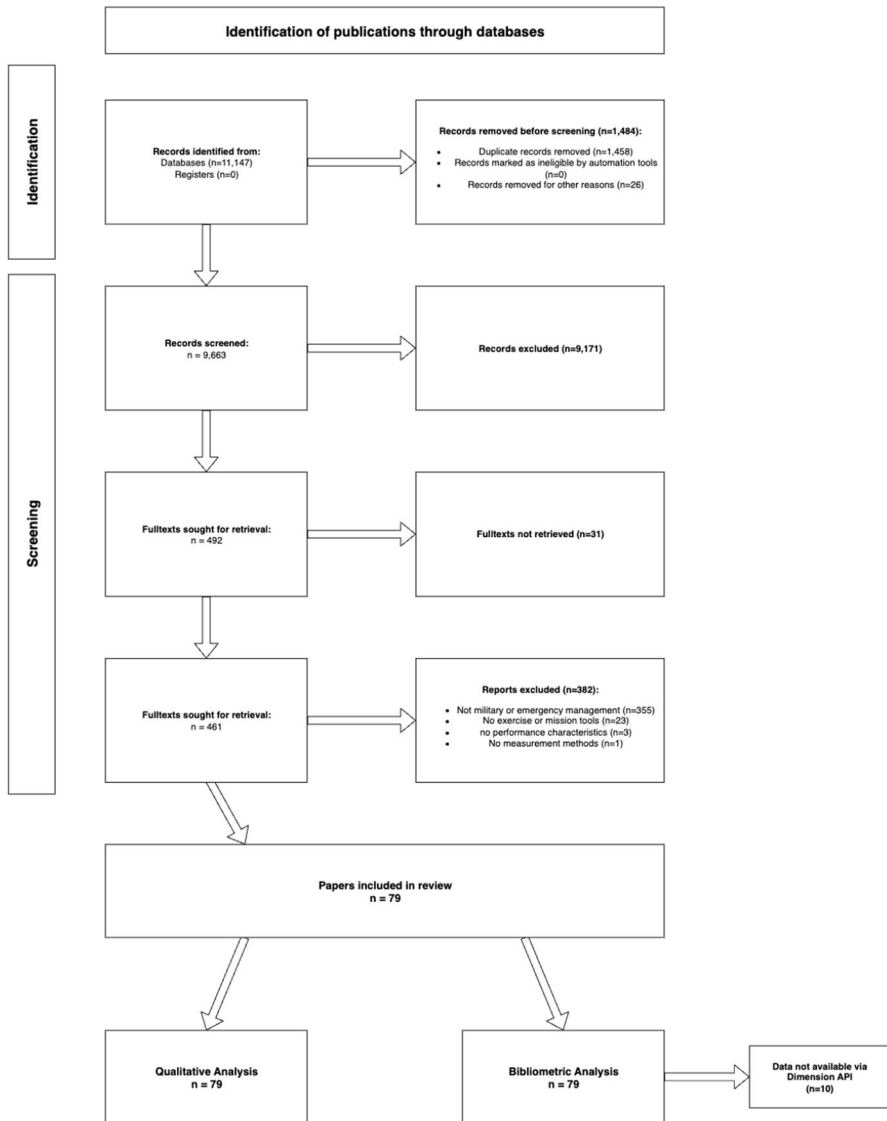
Fig. 1 PRISMA 2020 statement based on Page et al. (2021)

abstract screening resulted in 492 records that were selected for further review. Thirty-one articles could not be retrieved via library service and direct author contact and were excluded. The last step was a full text screening. This step of the selection process resulted in 79 records for the final analysis. Figure 1 shows the PRISMA 2020 statement (Page et al. 2021) with more detailed information about the excluded documents.

2.4 Stage 4: Charting the Data

The analysis was structured in two parts. First, a quantitative analysis of the 79 papers remaining for an in-depth review, and second, a qualitative assessment focusing on concepts of performance in command and control, methodological approaches and metrics for performance evaluation, and assessment approaches in exercises.

2.4.1 Quantitative Analysis

For the quantitative analysis, we utilized Digital Science's Dimensions Analytics API (Herzog et al. 2020) and the 'dimcli' (GitHub 2023) library in Python. The coverage of the Dimensions dataset on scientific production is comparable to other scientific databases (Harzing 2019). Therefore, it is suitable for the bibliometric analysis of this study. The integrated development and learning environment (IDE) for code development and execution was a 'Jupyter Notebook' (NumFOCUS 2023). The publication metadata (e.g., author, publication title, DOI, PMID, PID) were uploaded to a Jupyter environment. Publications lacking an identifier (DOI, PID, and PMID) had to be excluded. These were mostly gray literature works such as project reports from EU- or BMBF-funded projects. For publications with an identifier ($n = 69$), metadata stored within the Dimensions database was retrieved and further analyzed.

Units of assessment (UoA)and concepts: For the analysis of the subject area, the Units of Assessment (UAO) were used.

> "The Units of Assessment (UoA) are the 34 categories of research used by the Research Excellence Framework (REF) 2021 in the United Kingdom" (Digital Science & Research Solutions Inc. 2020).

The integration process into the Dimension database employs a machine learning algorithm. The concept extraction is also based on machine learning algorithms automatically extracting the

> "noun phrases [...] from a document's abstract as well as the rest of the Dimensions database [...]" (Digital Science & Research Solutions Inc. 2020).

2.4.2 Qualitative Synthesis

The qualitative assessment used the qualitative content analysis process of Mayring and Fenzl (2014). The whole process was facilitated using VERBI Software's MaxQDA version 2023, a tool specifically intended for qualitative research. All metadata and their corresponding full-text records were imported into the software for further examination. The coding process employed a hierarchical structure for the identified concepts and methods. The codes were derived on the basis of the three supporting research questions following an inductive
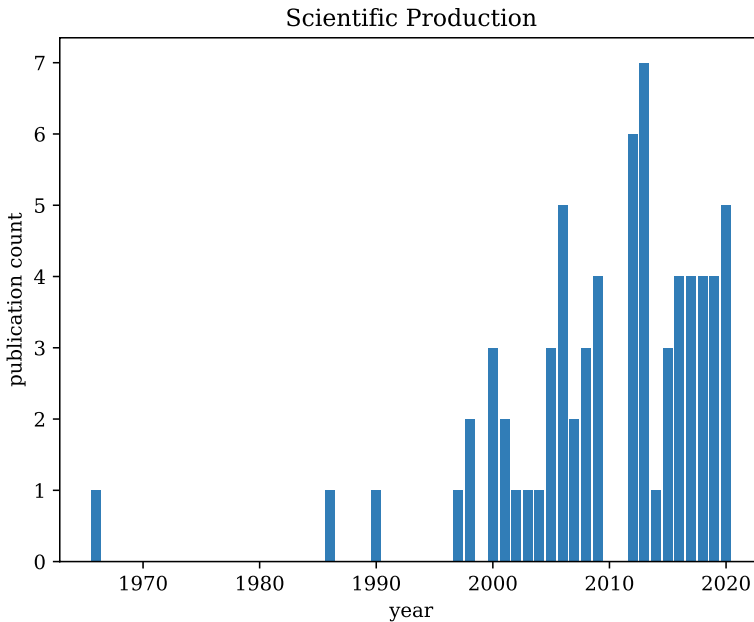
Fig. 2 Scientific production ($n = 69$) based on dimensions data

approach. After reviewing and coding all the papers, a second round was performed to check each paper for missing codes. After these two rounds, the coded text passages of each paper were summarized. These summaries were consolidated at the code level.

3 Results

3.1 Quantitative Analysis

The temporal distribution of the publications shows a raise of publications beginning at the end of the 1990s and a stable level until the study took the snapshot in 2021 (cf. Fig. 2). This could be an indication that the relevance of the field is increasing, although indexing scientific work elaborated over time and publications are increasing in general.

Regarding the research field the Units of Assessment (UoA) were used. Most publications came from business and management studies, followed by engineering (cf. Fig. 3). A significant part is from health-related fields (UoA: A03, A04, A02).

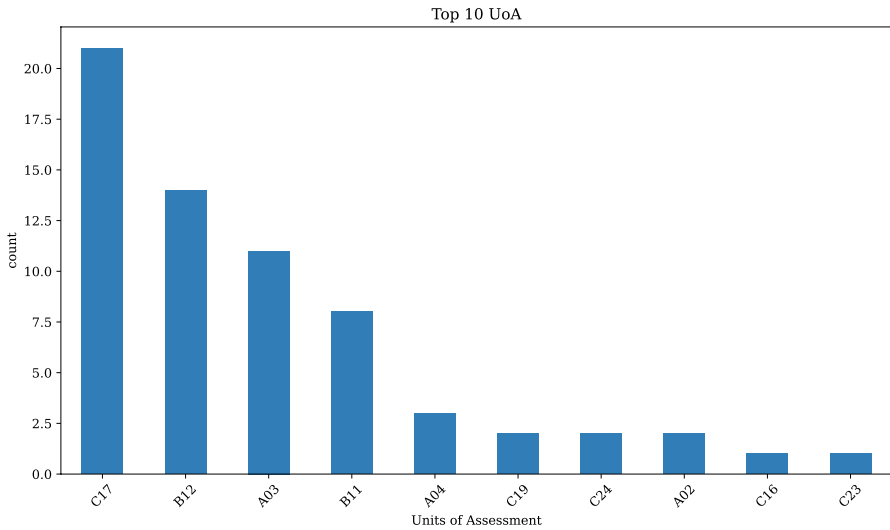Most productive authors were A. RÜters, N. A. Stanton, T. Vikström and H. Nilsson (cf. Fig. 4).

Fig. 3 Units of assessement (UoA) (*n* = 69) based on dimensions data (limitations refer to Sect. 4.1) UoA: C17–Business and Management Studies, A02–Public Health, Health Services and Primary Care, A03–Allied Health Professions, Dentistry, Nursing and Pharmacy, A04–Psychology, Psychiatry and Neuroscience, B11–Computer Science and Informatics, B12–Engineering, C16–Economics and Econometrics, C19–Politics and International Studies, C23–Education, C24–Sport and Exercise Sciences, Leisure and Tourism
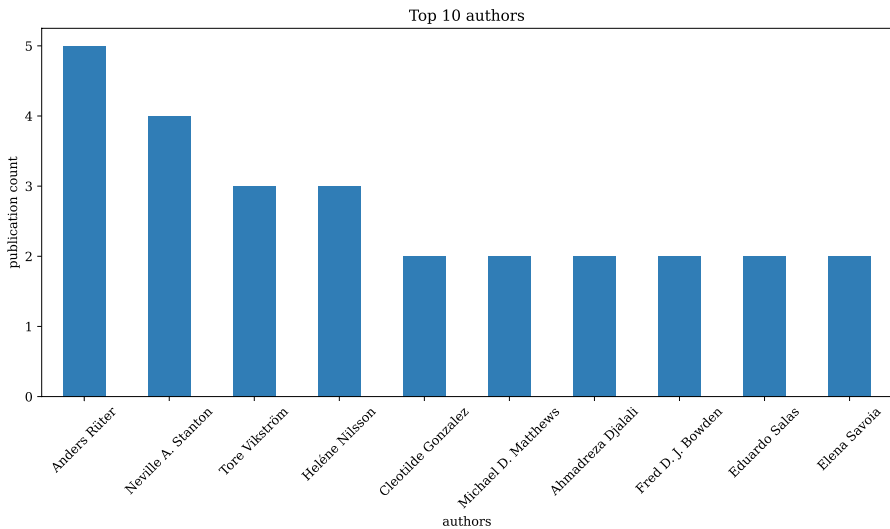

Fig. 4 Most productive authors (*n* = 69) based on dimensions data

The most active research organizations were Harvard University, Linköping University & University Hospital, and Texas A &M University (cf. Fig. 5). Research institutions in the United States published their results most frequently.
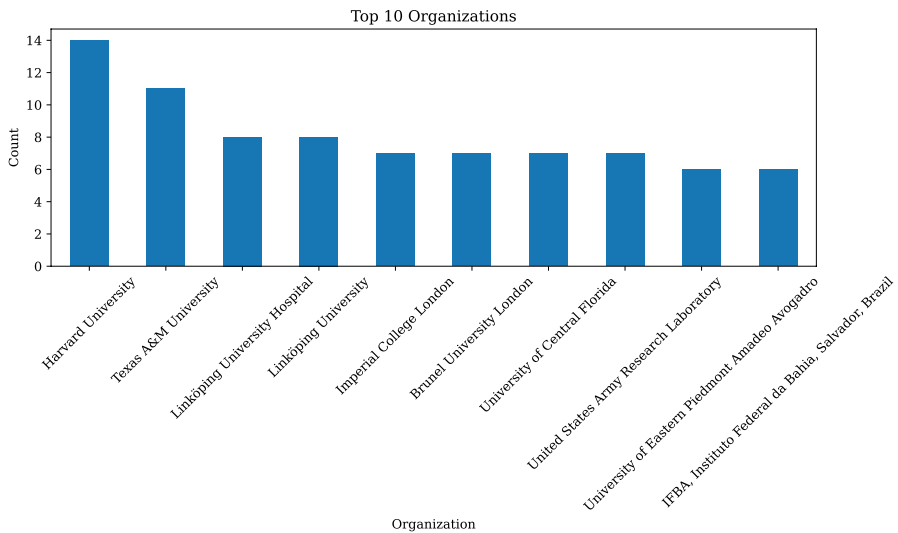
Fig. 5 Most active research organizations ($n = 69$) based on dimensions data
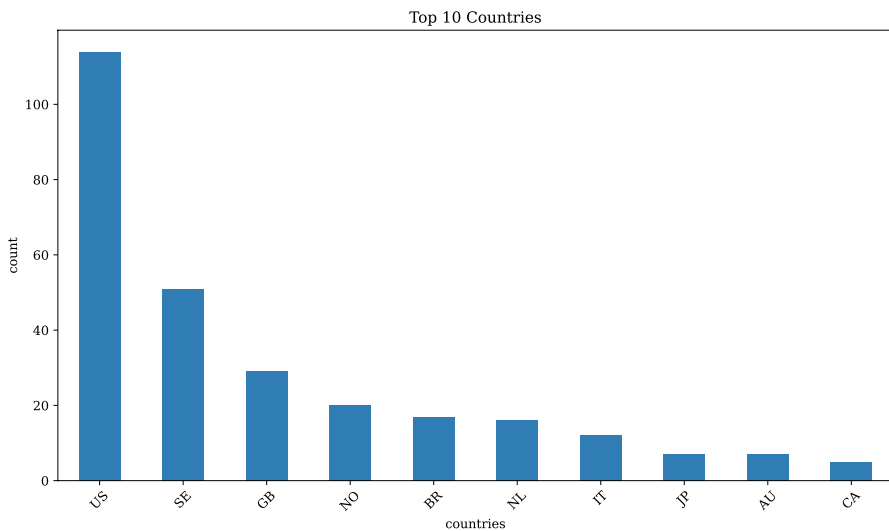

Fig. 6 Countries contributing ($n = 69$) based on dimension data

In a comparison of countries, U.S. researchers were most frequently involved, followed by the United Kingdom and Sweden (cf. Fig. 6).

Most of the studies focused on method development ($n = 25$), followed by examination of team processes ($n = 13$) and evaluation of teaching or training interventions ($n = 9$). Table 3 provides an overview.

Table 3 Background of study
(*n* = 79)

Background of study	*n*
Evaluation of concept/system/intervention	
(IT) System evaluation	4
Exercise evaluation	6
Teaching / training evaluation	9
Problem identification	5
Competence identification	2
Concept	
Team processes	13
Workload	2
Situational awareness	2
Communication	1
Development	
Method development	25
Framework developement	5
Description of a concept	2
Other	
Meta study	1
Thesis / dissertation	2
	79

Most of the studies were experimental studies (*n* = 33) followed by reviews of the literature (*n* = 11) and observational studies (*n* = 10). Table 4 lists these studies.

The study population consisted mainly of emergency management personnel (*n* = 35), followed by military personnel (*n* = 27). Six public health related studies focused on public health professionals while seven studies used students as study participants. Fahadullah's (2009) study lacked a clear statement (cf. Table 5).

3.2 Qualitative Synthesis

The qualitative synthesis focused on the three research questions presented in Sect. 2.1.

3.2.1 Performance Factors Influencing Command Post's Performance

Table 6 presents an overview of the factors presented in the following section.

Scenario: The scenario affects the performance of command posts (Worm 2001; Savoia et al. 2012; Stanton et al. 2015; Alavosius et al. 2017; Stadt Gelsenkirchen - Referat Feuerwehr 2019; To et al. 2019; Holdsworth and Zagorecki 2020). Both the initial state and the course of actions play a role. When exercises are used for learning purposes, the scenario has to support learning. If exercises are used for

Table 4 Method of study (n = 79)

Method	n	Studies
Literature review	11	Alavosius et al. (2017), Bearman et al. (2018), Bernier et al. (2012), Cosenzo et al. (2007), Heuverswyn and Huybrechts (2018), Kayman and Logar (2016), Qari et al. (2019), Salas et al. (2000), Savoia et al. (2012), Whitmore (2005), Wilson et al. (2007)
Delphi study	1	Gebbie et al. (2006)
Oberservational study	10	Cerqueira et al. (2017), Djalali et al. (2012), Djalali et al. (2014), Stadt Gelsenkirchen - Referat Feuerwehr (2019), Haar et al. (2017), Helsloot (2005), McCallum et al. (1990), Murphy et al. (2020), van Ruijven et al. (2015), Son et al. (2020)
Case study	9	Garvin and Miller (1981), Albinsson and Fransson (2001), Cohen et al. (2013), Gomes et al. (2014), Baroutsi (2016)
Surveying instrument[1]	3	Ceschi et al. (2019), Roud et al. (2021), Savoia et al. (2009)
Experimental study	33	Buchler et al. (2018), Dalenberg et al. (2009), Eid et al. (2005), Fahadullah (2009), Fiedler (1966), Helton et al. (2013), Johnston et al. (2013), Jøsok et al. (2019), Kobbeltvedt et al. (2005), Kozůbek (2017), Lambert (2004), Legemaate et al. (2012), Mao et al. (2016), Marusich et al. (2015), Nilsson and Rüter (2008), Nilsson et al. (2013), Peck et al. (2017), Pleban et al. (2002), Rådestad et al. (2012), Roberts et al. (2018), Rothrock et al. (2009), Rüter et al. (2007), Rüter and Vikstrom (2009), Salmon et al. (2007), Saetrevik and Kvamme (2019), Sjøvold and Nissestad (2020), Shattuck and Woods (1997), Silenas et al. (2008), Stanton and Harvey (2016), Taber et al. (2013), To et al. (2019), Walker et al. (2014)
Interview study[1]	3	Ceschi et al. (2019), Dixon et al. (2017), Riley et al. (2006), Sjöberg et al. (2006)
Other	8	Abraham (1986), Bouthonnier and Levis (1984), Gaertner et al. (2000), Reeves et al. (1998), Perry and Bowden (2003), Veenema et al. (2017), Worm (2001)
n	80[1]	

[1]The study of Ceschi et al. (2019) used more than one method

Table 5 Study population
($n = 79$)

Study population	n
Military personnel[1]	27
Emergency management[1]	35
Public health	6
Students	7
No population[2]	4
Not clear	1
	80 [1]

[1] One study covered both, military personnel and students

[2] Method development without a specific population and meta-studies

evaluation, there is a tension between the realistic course of an exercise and the conditions that can be controlled for (quasi)experimental purposes (Mendonca et al. 2006).

Time pressure is a significant factor that affects staff work. Missing time inevitably requires making assumptions and, subsequently, increases the need for changing and adapting plans more frequently (Riley et al. 2006). Time pressure does not automatically lead to a higher rate of procedural errors (Kobbeltvedt et al. 2005). In highly dynamic situations, personnel may not have the opportunity to make decisions but rather rely on decisions made by others (Abraham 1986; Helsloot 2005).

Workload is defined as the relationship between task demands and (perceived) available resources (Jacobs et al. 2013). According to Helton et al. (2013), workload contains both state and trait components. Workload has an impact on the error rate (Helsloot 2005; Alavosius et al. 2017). The influence is on the 'working memory' and affects contextual knowledge at the domain and strategy level (Gregoriades and Sutcliffe 2006). Causes of too much workload may be due to information sharing. However, communication increases with more demanding tasks. Roberts et al. (2018) describe a significant increase in the frequency of conversations in the context of submarine maneuvers, where more information is then transmitted, and a higher number of tasks are managed simultaneously. In principle, overload should be avoided as it can lead to stress reactions (Alavosius et al. 2017).

Resilience: If the command post is self-affected, the operational efficiency is compromised. The level of affectedness has a strong impact on performance (Helsloot 2005). Therefore, adaptability is a vital capability (Gomes et al. 2014) of the personnel, referring to their ability to adjust to changing circumstances and execute appropriate responses (Helsloot 2005; Whitmore 2005; Riley et al. 2006; Wilson et al. 2007; Haar 2014; Son et al. 2020). The availability and evaluation of information significantly impact the capacity to adapt. Stress-resistance is an important trait in emergency response fields in general (Cerqueira et al. 2017; To et al. 2019).

Table 6 Performance factors influencing command post's performance

Factor	Component	Description	Sources
Scenario		Initial state and course of actions of an exercise or a real incident	Worm (2001), Savoia et al. (2012), Stanton et al. (2015), Alavosius et al. (2017), Stadt Gelsenkirchen - Referat Feuerwehr (2019), To et al. (2019), Holdsworth and Zagorecki (2020), Mendonca et al. (2006)
	Time pressure	Time pressure arises from the rapid sequence of events or the pressure to make decisions	Riley et al. (2006), Kobbeltvedt et al. (2005), Abraham (1986), Helsloot (2005)
	Workload	Relationship between task demands and (perceived) available resources	Jacobs et al. (2013), Helton et al. (2013), Alavosius et al. (2017), Gregoriades and Sutcliffe (2006), Roberts et al. (2018)
Resilience		The ability to adapt to changing environments and situations, as well as the capacity to avoid or resist becoming affected by incidents	Cerqueira et al. (2017), Cosenzo et al. (2007), Haar (2014), Kayman and Logar (2016), Son et al. (2020), Wilson et al. (2007)
Situational awareness		The gathering, evaluation, and projection of the effects of all factors and conditions necessary to assess and manage a situation	Endsley (1995), Alavosius et al. (2017), Fahadullah (2009), Gomes et al. (2014), Pleban et al. (2002), Helsloot (2005), Holdsworth and Zagorecki (2020), Salas et al. (2000)
	Shared cognition	Process of deriving actions from situational information	Alavosius et al. (2017), Cerqueira et al. (2017), Dalenberg et al. (2009), Marusich et al. (2015), Shattuck and Woods (1997), Sjøvold and Nissestad (2020)
	Sensemaking	Time pressure arises from the rapid sequence of events or the pressure to make decisions	Shattuck and Woods (1997)
Decision-making		The process of making decisions	Alavosius et al. (2017), Kayman and Logar (2016), Kobbeltvedt et al. (2005), Shattuck and Woods (1997), Worm et al. (1998), Wilson et al. (2007)

Table 6 (continued)

Factor	Component	Description	Sources
Team structures and teamwork		The gathering, evaluation, and projection of the effects of all factors and conditions necessary to assess and manage a situation	Alavosius et al. (2017), Buchler et al. (2018)
	Coordination	Coordinating the team-members within the C2-Team	van Ruijven et al. (2015)
	Team cohesion	The ability of the team to work as a whole body	Salas et al. (2000), Alavosius et al. (2017), Cerqueira et al. (2017), Dalenberg et al. (2009), Marusich et al. (2015), Shattuck and Woods (1997), Sjøvold and Nissestad (2020)
	Staffing level	Team size and structure	Alavosius et al. (2017), Fiedler (1966), Gomes et al. (2014), Stanton et al. (2015), To et al. (2019)
	Personality of the Leader	Experience, characteristics and behavior of the leader	Baroutsi (2016), Salas et al. (2000), Shattuck and Woods (1997), To et al. (2019), Veenema et al. (2017)
	Experience	Experience of the team members in dealing with similar situations	Cerqueira et al. (2017), Stadt Gelsenkirchen - Referat Feuerwehr (2019), Helsloot (2005), Holdsworth and Zagorecki (2020), Savoia et al. (2012), Sjöberg et al. (2006), To et al. (2019), Wilson et al. (2007)
Operational execution		The technical implementation of the tasks to be performed	Baroutsi (2016), Stadt Gelsenkirchen - Referat Feuerwehr (2019), Gomes et al. (2014), Savoia et al. (2012)
	Information management	Process of deriving actions from situational information	Gomes et al. 2014, Helsloot (2005), Marusich et al. (2015), van Ruijven et al. (2015), Son et al. (2020), Stanton et al. (2015), Worm et al. (1998); Worm (2001)
	Communcation	Process of deriving actions from situational information	Dalenberg et al. (2009), van Ruijven et al. (2015), Rüter et al. (2007), Wilson et al. (2007)

Although stress depends on the individual's ability to cope with stress (Cosenzo et al. 2007), responding to stress can lead to cognitive biases. For example, fatigue may be recognized too late (Sjöberg et al. 2006). In this case, mechanisms must be established within command posts to recognize stress reactions and to have options available to reduce their effects (Kayman and Logar 2016).

Situational awareness (SA): SA is a key concept for staff performance (Shattuck and Woods 1997; Worm et al. 1998; Salas et al. 2000; Helsloot 2005; Riley et al. 2006; Alavosius et al. 2017; Holdsworth and Zagorecki 2020). The concept was originally developed by Endsley (1995). The construct consists on three levels of situational awareness: perception, comprehension, and projection. It has become the central concept in high-risk organizations (Alavosius et al. 2017). To produce SA, relevant, timely information about the state on the ground is essential. The production of SA is supported by the use of visualization elements such as geographic information systems (GIS) (Fahadullah 2009) or certain process elements such as timely, structured situation briefings (Gomes et al. 2014). In information processing, mainly terminologies and abbreviations are responsible for low SA values. A high SA value correlates with good values in decision-making (Pleban et al. 2002).

Within the team SA needs to be distributed to all team members. This distribution builds *shared cognition*. A mere exchange of information does not necessarily lead to better situational awareness (Marusich et al. 2015). The common goal and the overall picture of the situation must be understood and internalized by the team members (Shattuck and Woods 1997; Sjøvold and Nissestad 2020). The leader is responsible for ensuring common understanding (Alavosius et al. 2017). This can be achieved by performing team strategy discussions (Dalenberg et al. 2009) in which the shared picture of the goals, the situation, and the task distribution of the mission is developed. Another methodological approach is the development of a mental map (Cerqueira et al. 2017) with which knowledge, solution ideas and developments can then be represented.

Sensemaking is the process of deriving the necessary actions from situational information to achieve the overarching goal. In the case of 'anomalies,' the following six actions are relevant (a) seeking further information, (b) situation (self and damage situation), (c) referencing standard procedures, (d) referencing the incident commander's intent, (e) sequencing events, and (f) coordinating activities (Shattuck and Woods 1997). According to Dixon et al. (2017), in extreme situations where life is at risk, individuals may experience a flow state that facilitates rapid sensemaking.

Decision-making The command post is a decision-making body. The effectiveness of a command post depends on its ability to make decisions (Worm et al. 1998). Various factors can influence the decision-making process, including uncertainty (Shattuck and Woods 1997; Worm et al. 1998) ethics, cognitive biases, politics, moral principles, and personality traits of the staff members (Kayman and Logar 2016). The quality of decision-making is directly related to the availability of relevant information. In the studies reviewed (Kobbeltvedt et al. 2005; Wilson et al. 2007; Alavosius et al. 2017), sleep deprivation or fatigue was mentioned as a key

physical influence. Here the influence on decision making is mentioned. Kobbelt-vedt et al. (2005) found a reduction in speed with sleep deprivation, although accuracy remained the same. However, high-risk (life-threatening) motivation can mitigate performance-attenuating physical effects (Kobbeltvedt et al. 2005).

Team structures and teamwork Team structures impact performance. Objective evaluations show tuned-in teams achieve better results. Alavosius et al. (2017) notes that effective teamwork spans beyond communication to encompass individual member's behavior. Buchler et al. (2018) concur that increased face-to-face contact translates into decreased results in cyber operations. The 'right' amount of *coordination* is a crucial element of teamwork. 'Over-coordination' decreases performance (van Ruijven et al. 2015). According to Salas et al. (2000), *team cohesion* is achieved through common goals, performance monitoring, feedback, closed-loop communication, and support of others. These essential behaviors are considered to be necessary for effective teamwork. The *staffing level*, or team size, effects performance, requiring a balance between workload and increased coordination effort (Alavosius et al. 2017). Homogeneous groups may underperform compared to heterogeneous ones (Fiedler 1966; Gomes et al. 2014; To et al. 2019). Surprisingly, trust seems to have no influence on performance (Stanton et al. 2015).

The *personality* of the leader appears to have an impact on the performance of the command post team (Salas et al. 2000; Baroutsi 2016; Veenema et al. 2017; Jøsok et al. 2019). The behavior and characteristics of the leader need to be considered. These have an impact on different leadership styles and ways of working. Uncertainty tolerance appears to be an important trait for leaders. Unsuccessful leaders had low uncertainty tolerance and tended to be indecisive (Shattuck and Woods 1997). In addition to personal characteristics, expertise is also critical (To et al. 2019).

Moreover, the need for expertise in a given task or role is a key requirement (Helsloot 2005; Savoia et al. 2012; Cerqueira et al. 2017; Stadt Gelsenkirchen--Referat Feuerwehr 2019; To et al. 2019; Holdsworth and Zagorecki 2020; Silenas et al. 2008). The performance of the command post depends on this expertise. In addition to technical knowledge, understanding the characteristics of team members is essential. When knowledge of both the task and the team members is present, teams communicate better and achieve better results (Wilson et al. 2007). Knowledge of the job domain and its boundaries is also important. Mixing tasks or changing roles should be avoided (Sjöberg et al. 2006).

Experience In command post work, experience is an essential factor (Pleban et al. 2002; Sjöberg et al. 2006; Savoia et al. 2012; Haar 2014; To et al. 2019). Haar (2014) describes members as highly experienced specialists. Inexperience leads to poorer performance because it leads to underestimation of elements. In particular, experienced people are able to make more intuitive decisions because they automatically sort and evaluate the relevant facts without being aware about this process. Kayman and Logar (2016) refer to the intuitive decision making process. Expert ratings are superior to algorithms if the parameters aren't too complex, random or long-term to be estimated and the expert had been exposed to similar circumstances

relevant times. Veenema et al. (2017), however, state that pure expertise does not necessarily lead to better results. Rather, personality and especially self-efficacy play a role. Shared experiences help to strengthen the knowledge about the team members (Stanton et al. 2015).

Operational execution How command posts operate internally is called operational execution. Errors mainly arise from flaws in the operational procedures (Stadt Gelsenkirchen - Referat Feuerwehr 2019). 'Thematic vagabonding', is described as a negative influence on the staff's performance. Nevertheless, the leader's behavior also plays a role. If the leader encourages members to express their opinions more frequently, this could lead to more wandering (Baroutsi 2016).

Furthermore, adherence to standardized working methods, such as the Incident Command System (ICS) in the USA or Dienstvorschrift 100 (Ausschuss Feuerwehrangelegenheiten, Katastrophenschutz und zivile Verteidigung (AFKzV) 1999) in Germany, and technical proficiency in formatting and managing information within the team are essential (Savoia et al. 2012). The layout of the room can impact team communication, potentially in a supportive manner (Gomes et al. 2014).

Crew ressource management (CRM) behavior, also in command post work appears important (Alavosius et al. 2017) since 40 percent of errors occur in the decision-making or communication process (Stadt Gelsenkirchen - Referat Feuerwehr 2019). CRM behavior follows three phases: briefing, working phase, and debriefing (Alavosius et al. 2017). In the briefing phase the team setup (roles & responsibilities), the task, and objectives are being considered by the team-leader. The working phase follows the briefing and team members report progress and observations. If needed, work is interrupted and a debriefing meeting is beeing conducted. The aim of that meeting is to adjust the approach to fulfil the given task. The debriefing phase follows after task completion or significant events. Within this phase the work completed is reviewed and further tasks or improvements for further operations are discussed.

The leadership behavior of the leader, in contrast, has an impact on how the command post operates (Buchler et al. 2018).

Information management The availability of relevant information is essential to the work of staff. The processing and presentation of information is a core task of employees (Worm et al. 1998; Worm 2001; Gomes et al. 2014; Stanton et al. 2015). In particular, conflicting data and high dynamics have an impact on processing (Son et al. 2020). Uncertainty depends on the amount of information available (Perry and Bowden 2003). A customized presentation helps to maintain the right level of detail (Gomes et al. 2014). There are different views on the right amount of relevant information. van Ruijven et al. (2015) found no overload, while Marusich et al. (2015) stated that the amount of available information does not necessarily lead to better decisions. Interestingly, Helsloot (2005) noted that the use of liaison officers increases the pressure on employees.

Communication Communication both internally and with external forces appears to be an important part of team performance. Here, the allocation of internal resources

by the team leader and the communication among team members are important for performance (Dalenberg et al. 2009). In particular, the patterns of how teams communicate with each other have an impact on performance. Wilson et al. (2007) distinguishes three types: information sharing, phraseology, and closed-loop communication. This refers to getting the right information to the right person, choosing the right terminology and appropriate wording, as well as being complete while being concise. Closed-loop communication is important to ensure the proper transfer and understanding of information. Good communication will eventually lead to better situational awareness. In handover situations, special attention must be paid to proper communication; this is where most information loss occurs (Rüter et al. 2007). The centrality of the communication has an influence: van Ruijven et al. (2015) found that a higher amount of decentralized coordination which was defined as communication between team members does not lead into higher performance. On the other hand centralization lead into a bottlenecks hindering the information flow. In general, it is important to find the right level of communication.

3.2.2 Indicators of Command and Control Performance

The selection of measurement criteria in the reviewed papers was related to the evaluation purpose or study design. In terms of performance measurement, this selection must include several aspects or influencing factors. In performance evaluation, process, effectiveness, and efficiency are pertinent dimensions (Gaertner et al. 2000). Consequently, outcome parameters should be distinguished from process parameters (Bernier et al. 2012; Haar 2014). Outcome parameters encompass damage control (Fahadullah 2009) and error rates (Stadt Gelsenkirchen - Referat Feuerwehr 2019). However, identifying errors in staff work is challenging as they are often compensated and thus camouflaged (Stadt Gelsenkirchen - Referat Feuerwehr 2019). Process-oriented aspects pertain to the course of command post work and its elements (Sweet 1989; Gebbie et al. 2006; Dalenberg et al. 2009; Bearman et al. 2018). Table 7 provides an overview.

Processual criteria: For example, when considering *Crew Resource Management*, six relevant factors can be identified as being relevant: communication, situational awareness, decision making, teamwork, resource management within the team (mutual support), and leadership (Alavosius et al. 2017). In a study by Buchler et al. (2018), sociometric factors were used for evaluation. In general, the selection of criteria should be based on systems rather than individuals.

Coordination can be done by distributing and redistributing tasks according to the situation (Buchler et al. 2018). In the study by van Ruijven et al. (2015), coordination was mapped by measuring the density and centrality of a network. Stanton et al. (2015) considered in their study the intentions of the military leader. An 8-point Likert scale was used. The same study measured trust using a Likert scale.

Situational Awareness (SA) is described as a key factor in CRM. Lack of SA is one of the most common causes of aviation accidents (Ceschi et al. 2019). The three stages of SA perception, comprehension, and projection are the focus of interest

Table 7 Indicators of command and control performance

Criteria	Indicator	Sources
Processual criteria	Crew Resource Management (CRM)	Alavosius et al. (2017), Buchler et al. (2018)
	Coordination	Buchler et al. (2018), van Ruijven et al. (2015), Stanton et al. (2015)
	Situational awareness	Alavosius et al. (2017), Ceschi et al. (2019), Lambert (2004), Salmon et al. (2009), Taber et al. (2013)
	Shared mental models	Dalenberg et al. (2009), Sætrevik and Kvamme (2019)
	Sensemaking	Baroutsi (2016)
	Cooperation	Buchler et al. (2018), Ceschi et al. (2019), Savoia et al. (2012), To et al. (2019)
	Communication	Kozůbek (2017), Marusich et al. (2015), van Ruijven et al. (2015), Stanton et al. (2015)
	Leadership	Buchler et al. (2018), Ceschi et al. (2019), Haar et al. (2017)
	Decision-Making	Abraham (1986), Bernier et al. (2012), Ceschi et al. (2019), Haar (2014), Savoia et al. (2012)
	Sensemaking	Baroutsi (2016)
	Workload	Alavosius et al. (2017), Gregoriades and Sutcliffe (2006), Lambert (2004), Salmon et al. (2009), To et al. (2019), Whitmore (2005), Worm et al. (1998)
Outcome criteria	Effectiveness	Bernier et al. (2012)
	Efficiency	Bernier et al. (2012), Worm et al. (1998)
	Defined set of parameters	Bearman et al. (2018), Cohen et al. (2013), Djalali et al. (2014), Gaertner et al. (2000), Gebbie et al. (2006), Stadt Gelsenkirchen - Referat Feuerwehr (2019), Nilsson and Rüter (2008), Nilsson et al. (2013), Pleban et al. (2002), Rådestad et al. (2012), Rüter and Vikstrom (2009), Salmon et al. (2009), Walker et al. (2014), Whitmore (2005)

Table 8 Behavorial anchors for observing sensemaking (Baroutsi 2016)	Factor	Item [1]
	Resilience	Accelerating Feedback (+)
		Seek requisite variety (+)
		Self reflections (+)
		Normalizing (−)
		Increased risk behaviors (−)
	Goals	Hierarchical task structures (+)
		Goal elaborations (+)
		Confirmation seeking (−)
		Thematic vagabonding (−)
		Thematic lock-up (−)
	Expertise	Migrate discussions (+)
		Knowledge to reach expert (+)
		Ask for help (+)
		Fallacy of centrality (−)
		Avoiding decisions (−)
	Operations	Identify interruptions (+)
		Contact with the front-line (+)
		Seek out for bad news
	Errors	Detect small emerging failures (+)
		Specify unwanted mistakes (+)
		Clarify good news (+)
		Search strategy contradictions (+)
		Search incorrect assumptions (+)
	Don't simplify	Avoid categorization (+)
		Think/question out loud (+)
		Encourage different views (+)
		Notice unique aspects (+)
		Stereotypes/global hypothesis (−)

[1](+) represents behaviours which support sensemaking, (−) those which contradict sensemaking

(Lambert 2004; Salmon et al. 2009; Taber et al. 2013; Alavosius et al. 2017; Ceschi et al. 2019). In group situations, *shared mental models* are a concept of interest (Dalenberg et al. 2009; Sætrevik and Kvamme 2019). It describes the presence of the same shared understanding of a situation, goals, and necessary tasks.

Baroutsi (2016) presents behavioral anchors for observing *sensemaking*: These include 6 factors influencing sensemaking: resilience, goals, expertise, commitment, mistakes, and 'don't simplify'. Every factor consists of 3 to 5 items. Table 8 provides an overview.

Within teams, *cooperation* is an essential skill for effective teamwork (Savoia et al. 2012; To et al. 2019). It includes team building, mutual support, conflict resolution, task assignment, and situational reallocation (Buchler et al. 2018). In contrast

to leadership cooperation focuses more on the aspects of group climate and mutual support. The behavioral anchors here are: active and open communication, coordination of efforts, common goal setting, avoidance of personal differences (Ceschi et al. 2019). To establish cooperation, communication between team members is necessary.

Communication is an important aspect of performance. Several aspects are considered (Kozůbek 2017): communication behavior, information exchange, number of contacts and length of exchange (van Ruijven et al. 2015), frequency of information exchange, and content analysis (Marusich et al. 2015; Stanton et al. 2015).

Leadership is both a construct and a factor (Buchler et al. 2018). As construct leadership refers to the responsible management of people or an organization (Ceschi et al. 2019). Goal orientation is mentioned as an important aspect (Haar et al. 2017) as well as situation analysis and situational adaptation of the leadership style. Relevant behavioral anchors are using authority and assertiveness, maintaining standards, planning and coordination, managing workload and resources (Ceschi et al. 2019). Leadership as a factor is measured as part of different measuring scales (like the observational scaled assessment of teamwork—OAT).

A central construct in the context of leadership teams is *decision-making*. It is the inclusion and evaluation of all relevant information available at the time of the decision and the derivation of necessary actions (Haar 2014). Poor decision-making is involved in 47 percent of accidents (Ceschi et al. 2019). Decision-making processes vary depending on the time available. Possible supporting criteria are therefore the speed of the decision (Abraham 1986) and the quality in terms of relevance. The *level and mode of control* is a parameter for adaptation (Savoia et al. 2012). This refers to appropriate decisions compared to the leader's intentions, shared understanding, and shared values. Possible metrics are: elapsed decision-making period (time to reach a certain decision), number of alternatives considered, and number of alternative plans developed (Bernier et al. 2012).

Workload is both, an individual and team parameter (Worm et al. 1998; Lambert 2004; Salmon et al. 2009; Alavosius et al. 2017; To et al. 2019). It is understood as the management of the capacity of the individual members (Gregoriades and Sutcliffe 2006). The initial question is: 'Can the team/individual complete the tasks assigned to them within the time frame?' Workload can be an important parameter, as even highly trained personnel should not perform more than two tasks at the same time. As soon as knowledge-intensive tasks are increasing, the processing speed is reduced accordingly (Gregoriades and Sutcliffe 2006). Resource management is an alignment parameter according to Whitmore (2005).

Outcome criteria: The evaluation of the mission is usually done in terms of the outcome. One parameter is effectiveness (Bouthonnier and Levis 1984), i.e., the comparison between the target and the actual situation. Other parameters are efficiency, adaptability, flexibility and synergy. Bernier et al. (2012) see *effectiveness* as the ability to achieve the main goals. They introduce redundancy as another parameter that evaluates the number of 'nodes' (e.g. command posts) remaining in

the chain of command after an attack as well as the remaining links. Worm et al. (1998) describes efficiency as the relationship between the result and the resources used. Additionally, Bernier et al. (2012) define *efficiency* in terms of time, namely the fraction of time required for an operational response.

Once a *defined set of parameters* is used, an existing standard or common practice is used as a reference (Whitmore 2005; Walker et al. 2014; Salmon et al. 2009; Rüter and Vikstrom 2009; Rådestad et al. 2012; Pleban et al. 2002; Nilsson and Rüter 2008; Nilsson et al. 2013; Gebbie et al. 2006; Qari et al. 2019; Djalali et al. 2014; Cohen et al. 2013). The parameters include elements of the standard such as the conduct of meetings, the content of defense plans, information to higher and lower levels. In the ATHEBOS project (Stadt Gelsenkirchen--Referat Feuerwehr 2019), known common errors were used as parameters. Task completion is measured on a team or individual level. The quality (accuracy) of the tasks can also be included here (van Ruijven et al. 2015; Gaertner et al. 2000). Team outputs are effectiveness parameters. They are usually measured in terms of meeting target criteria (Bearman et al. 2018; Gaertner et al. 2000).

3.2.3 Evaluation of C2 Exercise Performance

When measuring performance in exercises, both the context and the scenario are important. Both parameters determine the course of the exercise and the corresponding responses. Process and outcome parameters should be collected either way. A whole methodology for setting up, conducting and debriefing of exercises is presented be the "e-notice" project (Heuverswyn and Huybrechts 2018). Salas et al. (2000) suggest four categories for evaluation, grouped in a four-field matrix: (1) team and (2) individual and (3) process and (4) outcome. Automated systems can help to better evaluate exercises and represent them in an event model (Holdsworth and Zagorecki 2020). When selecting parameters, issues of survey feasibility and cost must also be considered. Appropriate taxonomies for the task domain are also needed (Holdsworth and Zagorecki 2020). Graphical representations can help with interpretation, but they must be context sensitive. In addition, interpretation must take into account the context and embed the results in that context.

Measurement at the team level appears to have advantages over measurement at the individual level (Alavosius et al. 2017). In this context, monitoring tools are good for identifying problems but not for solving them. Methods used in SA assessments include *sample freezing* techniques, e.g., the Situation Awareness Global Assessment Technique (SAGAG) (Alavosius et al. 2017; Salmon et al. 2009; Taber et al. 2013). In addition, questionnaires, observers, or metrics are also being used. When measuring, care should be taken to maintain a good balance between accuracy and overload of the exercise participants (Mendonca et al. 2006).

A great part of the reviewed papers used observation and evaluationas main method. Here the behavior and the procedure are evaluated. Observation also uses experts. They observe the exercise and provide their expertise. During observation, observers may be part of the exercise and provide assistance or clarification or ask explicit questions. In other circumstances, they are silent participants who focus only on observations. Observations can be consensually summarized at the

end (Djalali et al. 2012). The observations focus on the object of investigation, e.g., team behavior or task completion. Tools such as questionnaires, checklists, or protocols (Cohen et al. 2013; Garvin and Miller 1981) augment the procedure for this purpose (Peck et al. 2017). These observations can be supported by behavioral anchors: Behavioral anchors are a common method supported by the use of checklists or rating scales (Bearman et al. 2018) such as the Air Warfare Team Performance Index (APTI) (Johnston et al. 2013; Reeves et al. 1998). In some cases, behavioral anchors are used as performance indicators of team behavior (Ceschi et al. 2019). The team's behavior is evaluated against a predefined ideal state (Alavosius et al. 2017). The analysis refers to the occurrence of this behavior.

Questionnaires as the sole survey instrument are usually used for *self-assessment* (Mendonca et al. 2006; Savoia et al. 2012; Worm 2001; Bearman et al. 2018; Roud et al. 2021; Eid et al. 2005; Legemaate et al. 2012) of teams. In addition, questionnaires can be used for team evaluation. In the event of an incident, questionnaires can provide valuable information about teamwork and behavior (Salas et al. 2000). This can be done as part of the exercise or immediately after the simulation. Questionnaire instruments can also support observers, e.g., the Disaster Management Indicator Scale (DiMI) (Murphy et al. 2020; Nilsson and Rüter 2008; Rådestad et al. 2012; Rüter et al. 2007), or the Observational Assessment of Teamwork (OAT) (Buchler et al. 2018). Questionnaires can be used to inquire about the behavior of other participants. There is a risk, especially with participant questionnaires, that they will affect the course of the exercise and the perceived realism. Established survey instruments that work with scales typically use rational, nominal, or ordinal scales, usually Likert scales.

Numerical methods refer to the measurement of parameters that can be quantified. For example, the number of messages exchanged (Fahadullah 2009; McCallum et al. 1990), the duration between two points in time (e.g., input and output, request and fulfillment, etc.) (Salmon et al. 2007), or the communication in terms of sentence length, word count, etc. (Bearman et al. 2018) are measured.

When measuring *task completion*, the number of completed tasks (van Ruijven et al. 2015) is measured as a value or set relative to the total number of tasks (Buchler et al. 2018; Rothrock et al. 2009). The quality of the completed tasks, such as compliance with predefined standards or requirements, can also be evaluated (Gaertner et al. 2000). It is reasonable to relate the number of tasks and quality to the time required.

Precision is the measurement of the performance of the result against an idealized standard. Here, error rates (deviation from the standard) (Kobbeltvedt et al. 2005; Mao et al. 2016) or the 'correct' representation of the situation in situation reports (Buchler et al. 2018) can be used as parameters.

Benchmarks exist with established standards for processing events. *Time* is a commonly used parameter. It measures the time it takes to reach certain milestones or to complete certain tasks (Alavosius et al. 2017; Albinsson and Fransson 2001;

Gaertner et al. 2000). When presenting the results of an exercise, a time axis is useful to put the processes of the exercise in order (Albinsson and Fransson 2001).

With the help of *sensors*, different data from the used systems (Buchler et al. 2018; Reeves et al. 1998) can be used to measure, e.g., temporal aspects or communication (Mendonca et al. 2006). *Biomarkers* are used to measure stress levels. Here, catecholamine or amylase levels are considered suitable surrogate parameters (Cosenzo et al. 2007; Worm 2001).

To compensate for the disadvantages of individual methods, a **mixed method approach** (Bearman et al. 2018; Holdsworth and Zagorecki 2020; van Ruijven et al. 2015) can be used, which may include, for example, technical parameters from computer systems and observation, or observation and questionnaires (Alavosius et al. 2017). In general, mixed-method approaches seem to be underrepresented in application. An example for implementing more than one method is the study of Mendonca et al. (2006). Here, the effectiveness of decisions was measured. A wider range of methods was used: time difference between insertion and decision, and degree of 'correctness' of the decision. After the simulation, the participants filled out questionnaires.

4 Discussion

Command posts are important parts of the disaster response system. Their functioning is essential for coordination and, in the end, for saving lives. This study explored scenario, resilience, situational awareness, decision-making, team structures, and teamwork, as well as operational execution, as factors influencing command post performance.

Compared to Heumüller et al. (2014), who introduced a conceptional model of command post exercises with the aim of guiding the conception of command post exercises, this study obtained similar results. The elements of Heumüller et al's model, namely *'management process'* ("Führungsprozess"), *'coordination areas'* ("Koordinationsbereiche"), *'internal structure'* ("Stabsstruktur"), and *'standard procedures'* ("Standardprozeduren") are similar to the *influencing factors* (Sect. 3.2.1) found in this study. In detail: The management process refers to decision making (Sect. 3.2.1, situational awareness (Sect. 3.2.1), since Heumüller et al's management process covers the decision making process of German DV100 (Ausschuss Feuerwehrangelegenheiten, Katastrophenschutz und zivile Verteidigung (AFKzV) 1999). Decision making is based on situational awareness whereas the shared understanding concept is closely related to situational awareness (Endsley 2000). Coordination areas and internal structure refer to team structures and teamwork (Sect. 3.2.1). Lastly, standard procedures are described as operational execution (Sect. 3.2.1) within this study. Furthermore, Heumüller et al described a reciprocal impact between the scenario and the command post model. While the scenario is being played out in an exercise, in real operations, it is dictated by the situation.

Gißler (2019) describes the performance of a command and control team as sufficient leadership performance. The team establishes the necessary conditions for a successful mission carried out by operational units through their own efforts.

A C2 team is, in his perspective, a 'Swiss army knife' for finding a solution to a mostly unknown or difficult to handle problem. This view leads to the request for a functioning unit that can only be measured by its way of solving the challenges given by the situation.

The key work of a C2-Team is preparing and enabling a decision. Interestingly, Kayman and Logar (2016) refer in this context to the concept of natural decision making (Kahneman and Klein 2009). Taking this concept into account, the decision-making process within the C2-Team could be biased or accelerated by intuitive expert judgments.

At this point, this study adds criteria for a deeper view into the work of a command post. In performance measurement, three elements—process, effectiveness, and efficiency—are suitable as evaluation levels. First, the outcome and process parameters should be distinguished: The outcome is the result of a number of processes within the command team. Stadt Gelsenkirchen - Referat Feuerwehr (2019); Gißler (2019), both, described that the outcome of the mission is not necessarily the result of the command post's work. It is difficult to identify errors in staff work by their effects because they are compensated and thus hidden. For this reason, the outcome evaluation should refer solely to the command post and not to the entire operation.

Process aspects are related to the course of staff work itself and the elements within it. The results of this study indicate that the processes in the command post depend on several aspects which are covered by crew resource management (CRM). CRM was originally developed for small teams, such as flight crews, in operational contexts. It became more popular for midsize teams, such as resuscitation teams or control room teams in power plants. Therefore, it seems appropriate to consider CRM in the context of command posts.

The observation and evaluation method is, compared to the work of Beerens and Tehler (2016), still the most common method for evaluating exercises. In principle, observations appear to be a cost-effective means of evaluating exercises. However, the cost-benefit ratio is questionable: Especially if reliable results are to be achieved, more time for training and preparation must be calculated. Nontheless, other methods can augment the evaluation giving a different view on the exercise. The use of software to support the work, such as GIS or just email for message exchange, can help to gather more information about processing times, for example. This can help observers to catch another view on the exercise and confirm or disprove their judgment.

Research in this context is difficult because research interests, such as controllable conditions, often lead to unrealistic exercise scenarios (Gomes et al. 2014). Several studies investigated command post performance with a specific focus on, for example, output, team processes, or SA. However, as a complex system, command posts do not rely on a single factor that influences performance but rather on a multitude of interdependent factors. Haar et al. (2017) tried to compensate that issue still using an observer method for data collection. Figure 7 presents an overview on dependencies derived from the findings of this study. This issue was tried to compensate by the work of Haar et al. (2017) still using an observer method for data collection.
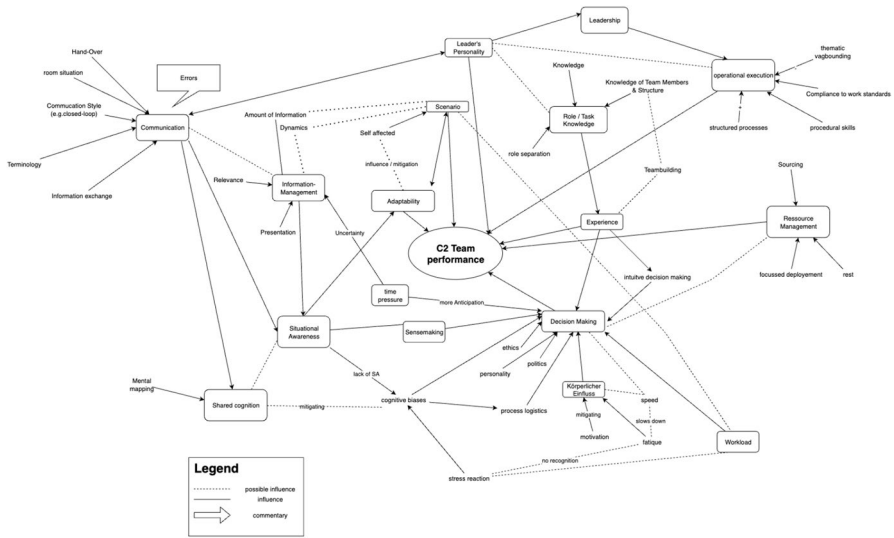
Fig. 7 Factors of performance and dependencies

Recent findings of Baroutsi (2023) analyzing 55 studies on quantitative performance evaluations show that research in this field is still necessary. Research methods for different situations are lacking. There are no empirical baselines, which hinders benchmarking of performance. This might be due to challenges in statistical validation. Baroutsi highlights the knowledge of specific emergency management doctrines to improve inter-rater reliability.

In conclusion, Beerens (2019), Beerens et al. (2020) established the need for a framework for evaluating emergency management exercises. Future research should focus on various assessment methods for these exercises, which is supported by the results of this study.

4.1 Limitations

In this study, several limitations have to be considered. First and foremost, the study covers papers until May 2021. In the meantime, more recent findings could have been published. In the discussion section, some newer publications known by the authors trying to mitigate these effects. The underlying research for this paper was done by the main author. This may lead to biases that were not uncovered through a second look at the results. A strict protocol (Nordhausen and Hirt 2020) was used to avoid these problems. Within the bibliometric analysis, not all papers were included due to missing identifiers. This issue concerns reports from research projects and papers that were not published in a journal article. In the bibliometric analysis the UoA were used for retrieving the research field of the paper. This may lead to a limitation because

"as with all of the categorisation systems in Dimensions, the UoA system is an algorithm-based model of the Units of Assessment, created through machine learning on pre-labeled documents. The UoA categorisation system in Dimensions is not a record-for-record copy of the UoA classifications, but a very accurate emulation. This means that there may be a small number of cases in which a publication submitted to the REF process was defined as being in one category, but when found in Dimensions it is labelled in a different category, or in more than one category." (Digital Science & Research Solutions Inc. 2020)

This study focused on a specific part, namely command posts in emergency management. Generalization of the results for emergency management itself is not possible, even if different areas have the same coverage.

5 Conclusion

This study examined the performance of command posts using a scoping review method to determine current knowledge on the subject. Six performance factors were identified: scenario, resilience, situational awareness, decision making, team structures and teamwork, and operational execution. When evaluating performance, attention should be paid to three dimensions: process, effectiveness, and efficiency. Teamwork is predominant within the process dimension. Therefore, performance is defined by CRM-related aspects. Effectiveness and efficiency are dimension that describe the outcome. Effectiveness focuses on achieving the desired outcome, e.g., if the necessary measures were taken. Efficiency is primarily determined by the time required to perform an action. Exercises often lack realism because the scenario is set to evaluate the system as a whole. Exercise planning and execution are time-consuming and cost-consuming. Not surprisingly, the number of exercises is small and exercise planners feel the pressure to cover as many agencies and organizations as possible. Consequently, they face challenges in designing exercises that effectively train or evaluate important aspects. However, this study allows exercise planners to focus on the pertinent aspects of C2 work and create suitable scenarios. The findings of this study have several practical implications. A potential application of the results would be to support scholars and practitioners in developing benchmarks and evaluation methods on C2 performance in a training and evaluation setting. The results found here can be used to better target training measures for C2-teams. The evaluation of training exercises follows an indicator-based approach, allowing for a more precise targeting of the training. In general, this approach can improve the performance of participants. In general, this can improve leadership performance in crisis management and reduce the costs associated with inadequate training measures. In addition, the results can lead in the creation of registers that compare the performance of management teams, allowing general conclusions to be drawn about team performance. These registers provide researchers and practitioners with information to improve crisis management as a whole. The complexity of command and control research is demonstrated by the fact that each performance factor that

influences command post performance is part of its own research field. Therefore, evaluating command and control performance remains challenging due to the various interdependencies among the factors identified in this study. Consequently, more research on this interdependence . A set of performance indicators could be useful in determining which factors lead to which outcome. This could help bridge the gap between the required realism and the scientific validity.

Declarations

Conflict of Interest Authors declare no financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

References

Abraham D (1986) System concept and criteria for battle group decision making. Naval Eng J 98(3):138–141. https://doi.org/10.1111/j.1559-3584.1986.tb03424.x

Alavosius MP, Houmanfar RA, Anbro SJ et al (2017) Leadership and crew resource management in high-reliability organizations: a competency framework for measuring behaviors. J Organiz Behavior Manag 37(2):1–29. https://doi.org/10.1080/01608061.2017.1325825

Albinsson PA, Fransson J (2001) Communication visualization–an aid to military command and control evaluation. Proc Human Factors Ergonom Soc Ann Meeting 45(6):590–594. https://doi.org/10.1177/154193120104500603

Arksey H, O'Malley L (2005) Scoping studies: towards a methodological framework. Int J Soc Res Methodol 8(1):19–32. https://doi.org/10.1080/1364557032000119616

Ausschuss Feuerwehrangelegenheiten, Katastrophenschutz und zivile Verteidigung (AFKzV) (1999) Feuerwehr-Dienstvorschrift 100: Führung und Leitung im Einsatz - Führungssystem. AK V der Innenministerkonferenz (IMK), https://www.bbk.bund.de/SharedDocs/Downloads/BBK/DE/FIS/DownloadsRechtundVorschriften/Volltext_Fw_Dv/FwDV%20100.pdf?__blob=publicationFile

Baroutsi N (2016) Observing sensemaking in C2: performance assessments in multi-organizational crisis response. In: Proceedings of the ISCRAM 2016 conference, http://idl.iscram.org/files/nicolettabaroutsi/2016/1385_NicolettaBaroutsi2016.pdf

Baroutsi N (2023) Quantitative performance evaluations for crisis response management: a scoping study. Int J Emergency Manage 18(2):107–143. https://doi.org/10.1504/ijem.2023.131932

Bearman C, Rainbird S, Brooks BP et al (2018) A literature review of methods for providing enhanced operational oversight of teams in emergency management. Int J Emergency Manage 14(3):254. https://doi.org/10.1504/ijem.2018.094237

Beerens RJJ (2019) Does the means achieve an end? A document analysis providing an overview of emergency and crisis management evaluation practice in the Netherlands. Int J Emergency Manage 15(3):221–254. https://doi.org/10.1504/ijem.2019.102310

Beerens RJJ, Tehler H (2016) Scoping the field of disaster exercise evaluation–A literature overview and analysis. Int Jo Disaster Risk Reducti 19:413–446. https://doi.org/10.1016/j.ijdrr.2016.09.001

Beerens RJJ, Tehler H, Pelzer B (2020) How can we make disaster management evaluations more useful? an empirical study of Dutch exercise evaluations. Int J Disaster Risk Sci. https://doi.org/10.1007/s13753-020-00286-7

Bernier M, Leblanc S, Morton B (2012) Metrics framework of cyber operations on command and control. In: Proceedings of the 11th European conference on information warfare & security : ECIW 2012, pp 53–62

Bouthonnier V, Levis AH (1984) Effectiveness analysis of C3-systems. IEEE Trans Syst Man Cybern SMC–14(1):48–54. https://doi.org/10.1109/tsmc.1984.6313268

Buchler N, Rajivan P, Marusich LR et al (2018) Sociometrics and observational assessment of teaming and leadership in a cyber security defense competition. Comput Security 73:114–136. https://doi.org/10.1016/j.cose.2017.10.013

Cerqueira I, Filho SA, Silva AJN, et al (2017) Analysis of competencies for decision crisis management in actions under progressive stress: a simulated case. Adv Intell Syst Comput pp 269–277. https://doi.org/10.1007/978-3-319-60525-8_28

Ceschi A, Costantini A, Zagarese V et al (2019) The NOTECHS+: A short scale designed for assessing the non-technical skills (and more) in the aviation and the emergency personnel. Front Psychol 10:902. https://doi.org/10.3389/fpsyg.2019.00902

Cohen D, Sevdalis N, Patel V et al (2013) Tactical and operational response to major incidents: Feasibility and reliability of skills assessment using novel virtual environments. Resuscitation 84(7):992–998. https://doi.org/10.1016/j.resuscitation.2012.12.011

Colquhoun HL, Levac D, O'Brien KK et al (2014) Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol 67(12):1291–1294. https://doi.org/10.1016/j.jclinepi.2014.03.013

Cosenzo KA, Fatkin LT, Patton DJ (2007) Ready or not: enhancing operational effectiveness through use of readiness measures. Aviat Space Environ Med 78(5 Suppl):B96-106

Dalenberg S, Vogelaar ALW, Beersma B (2009) The effect of a team strategy discussion on military team performance. Military Psychol 21(sup2):S31–S46. https://doi.org/10.1080/08995600903249107

Digital Science & Research Solutions Inc. (2020) What is the background behind the Units of Assessment categorisation system in Dimensions? : Dimensions. https://dimensions.freshdesk.com/support/solutions/articles/23000019506-what-is-the-background-behind-the-units-of-assessment-categorisation-system-in-dimensions-

DIN EN ISO 22300 (2021) Sicherheit und Resilienz - Begriffe (ISO 22300:2021); Deutsche Fassung EN ISO 22300:2021

Dixon DP, Weeks M, Boland R et al (2017) Making sense when it matters most: An exploratory study of leadership in extremis. J Leader Organiz Stud 24(3):294–317. https://doi.org/10.1177/1548051816679356

Djalali A, Castren M, Hosseinijenab V et al (2012) Hospital incident command system (HICS) performance in Iran; decision making during disasters. Scandinavian J Trauma Resuscit Emerg Med 20(1):14. https://doi.org/10.1186/1757-7241-20-14

Djalali A, Carenzo L, Ragazzoni L et al (2014) Does hospital disaster preparedness predict response performance during a full-scale exercise? A pilot study. Prehospital Disaster Med 29(5):441–447. https://doi.org/10.1017/s1049023x1400082x

Eid J, Meland NT, Matthews MD et al (2005) Dispositional optimism and self-assessed situation awareness in a Norwegian military training exercise. Percept Mot Skills 100(3):649–658. https://doi.org/10.2466/pms.100.3.649-658

Endsley MR (1995) Toward a theory of situation awareness in dynamic systems. Human Factors J Human Factors Ergonom Soc 37(1):32–64. https://doi.org/10.1518/001872095779049543

Endsley MR (2000) Situation models: An avenue to the modeling of mental models. Proc Human Factors Ergonom Soc Ann Meeting 44(1):61–64. https://doi.org/10.1177/154193120004400117

Fahadullah MNJM (2009) Statistical analysis of team training in emergency management simulator system. PhD thesis, Linköpings universitet

Fiedler FE (1966) The effect of leadership and cultural heterogeneity on group performance: A test of the contingency model. J Exp Soc Psychol 2(3):237–264. https://doi.org/10.1016/0022-1031(66)90082-5

Gaertner PS, Slade M, Bowden F et al (2000) Battlespace awareness and the Australian Army battlefield command support system. Digitiz Battlespace V Battlefield Biomed Technol II:23–34. https://doi.org/10.1117/12.395054

Garvin JM, Miller KP (1981) The logistics of a disaster drill: Medex-78. Emergency Med Serv 10(1):39–48

Gebbie KM, Valas J, Merrill J et al (2006) Role of exercises and drills in the evaluation of public health in emergency response. Prehospital Disaster Med 21(3):173–182. https://doi.org/10.1017/s1049023x00003642

GitHub (2023) digital-science/dimcli: Python client and CLI for scientometrics and research analytics using the Dimensions API. Version 1.1. https://github.com/digital-science/dimcli

Gißler D (2019) Erfolg der Stabsarbeit. Verlag für Polizeiwissenschaft, Frankfurt a. M

Gomes JO, Borges MR, Huber GJ et al (2014) Analysis of the resilience of team performance during a nuclear emergency response exercise. Appl Ergonom 45(3):780–788. https://doi.org/10.1016/j.apergo.2013.10.009

Gregoriades A, Sutcliffe AG (2006) Automated assistance for human factors analysis in complex systems. Ergonomics 49(12–13):1265–1287. https://doi.org/10.1080/00140130600612721

Haar Svd (2014) Getting on the same page : team learning and team cognition in emergency management command-and-control teams. PLATO : centre for research and development in education and training, faculty of social and behavioural sciences, Leiden University, Leiden, https://hdl.handle.net/1887/23623

Haar Svd, Koeslag-Kreunen M, Euwe E et al (2017) Team leader structuring for team effectiveness and team learning in command-and-control teams. Small Group Res 48(2):215–248. https://doi.org/10.1177/1046496417689897

Harzing A-W (2019) Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar Microsoft Academic Scopus and the Web of Science? Scientometrics 120(1):341–349. https://doi.org/10.1007/s11192-019-03114-y

Helsloot I (2005) Bordering on reality: findings on the bonfire crisis management simulation. J Contingencies Crisis Manag 13(3):159–169. https://doi.org/10.1111/j.1468-5973.2005.00472.x

Helton WS, Funke GJ, Knott BA (2013) Measuring workload in collaborative contexts. Human Factors J Human Fact Ergonom Soc 56(2):322–332. https://doi.org/10.1177/0018720813490727

Herzog C, Hook D, Konkiel S (2020) Dimensions: bringing down barriers between scientometricians and data. Q Sci Stud 1(1):387–395. https://doi.org/10.1162/qss_a_00020

Heumüller E, Richter S, Lechner U (2014) Ein konzeptionelles stabsmodell für die entwicklung von stabsrahmenübungen. In: Kundisch D, Suhl L, Beckmann L (eds) MKWI 2014 - Multikonferenz Wirtschaftsinformatik. Tagungsband, Paderborn, p 321-334, http://www.mkwi2014.de/Content/Tagungsband_MKWI2014.pdf

Heuverswyn KV, Huybrechts I (2018) D4.1 eNOTICE methodology for the preparation, organisation, evaluation and follow up of CBRN exercises combined with tests, validations or demonstrations. Technical Report, eNOTICE consortium, https://cordis.europa.eu/project/id/662284/reporting

Holdsworth D, Zagorecki A (2020) The SERIES model: development of a practitioner focused emergency response evaluation system. Int J Emerg Serv 9(3):313–337. https://doi.org/10.1108/ijes-01-2020-0003

Hufschmidt G, Dikau R (2013) Bildung als katastrophenvorsorge. In: Mitschke T, Unger C, Freudenberg D (eds) Krisenmanagement - Notfallplanung - Bevölkerungsschutz. Festschrift anlässlich 60 Jahre Ausbildung im Bevölkerungsschutz. Krisenmanagement - Notfallübung - Bevölkerungsschutz. Festschrift anlässlich 60 Jahre Ausbildung im Bevölkerungsschutz, Duncker & Humblot, Berlin, p 273–291, https://www.kavoma.de/Hufschmidt_Dikau_Bildung_als%20KatVorsorge_Festschrift.pdf

Jacobs K, Hellman M, Markowitz J et al (2013) Workload. In: Gellman MD, Turner JR (eds) Encyclopedia of behavioral medicine. Springer New York, New York, NY, pp 2068–2069. https://doi.org/10.1007/978-1-4419-1005-9_934

Johnston JH, Fiore SM, Paris C et al (2013) Application of cognitive load theory to develop a measure of team cognitive efficiency. Military Psychol 25(3):252–265. https://doi.org/10.1037/h0094967

Jøsok O, Lugo R, Knox BJ et al (2019) Self-regulation and cognitive agility in cyber operations. Frontiers Psychol 10:875. https://doi.org/10.3389/fpsyg.2019.00875

Kahneman D, Klein G (2009) Conditions for intuitive expertise: A failure to disagree. Am Psychol 64(6):515–526. https://doi.org/10.1037/a0016755

Kayman H, Logar T (2016) A framework for training public health practitioners in crisis decision-making. Disaster Med Public Health Preparedness 10(1):165–173. https://doi.org/10.1017/dmp.2015.149

Kobbeltvedt T, Brun W, Laberg JC (2005) Cognitive processes in planning and judgements under sleep deprivation and time pressure. Organ Behav Hum Decis Process 98(1):1–14. https://doi.org/10.1016/j.obhdp.2005.05.002

Kozůbek J (2017) Possibilities of leadership and tactical training evaluation. In: ICLEL 2017 Proceedings Book, vol 5. ICLEL conferences, Sakarya, pp 1092–1100, https://www.publication-iclel.com/article.php?page=possibilities-of-leadership-and-tactical-training-evaluation

Lambert NJ (2004) Analysis and evaluation of the command and control concept of the NATO immediate reaction task force (Land). J Oper Res Soc 55(4):375–389. https://doi.org/10.1057/palgrave.jors.2601666

Lamers C (2016) Führungssysteme bei komplexen Einsatzlagen. Forschungsbericht p 40. https://www.vfdb.de/fileadmin/download/bericht_prakos_2016-06-03_.pdf

Legemaate GAG, Burkle FM, Bierens JJLM (2012) The evaluation of research methods during disaster exercises: applicability for improving disaster health management. Prehospital Disaster Med 27(1):18–26. https://doi.org/10.1017/s1049023x11006789. http://gateway.webofknowledge.com/gateway/Gateway.cgi?GWVersion=2 &SrcAuth=SwissAcademicSoftware &SrcApp=Citavi &DestLinkType=FullRecord &DestApp=WOS &KeyUT=000445123100004

Levac D, Colquhoun H, O'Brien KK (2010) Scoping studies: advancing the methodology. Implement Sci 5(1):69. https://doi.org/10.1186/1748-5908-5-69

Mao A, Mason W, Suri S et al (2016) An experimental study of team size and performance on a complex task. PLoS ONE 11(4):e0153048. https://doi.org/10.1371/journal.pone.0153048

Marusich LR, Bakdash JZ, Onal E et al (2015) Effects of information availability on command-and-control decision making. Human Factors J Human Factors Ergonom Soc 58(2):301–321. https://doi.org/10.1177/0018720815619515

Mayring P, Fenzl T (2014) Qualitative inhaltsanalyse. In: Baur N, Blasius J (eds) Handbuch Methoden der empirischen Sozialforschung. Springer VS, Wiesbaden, pp 543–556. https://doi.org/10.1007/978-3-531-18939-0_38

McCallum MC, Bittner AC, Badalamente RV (1990) Empirical identification of user information requirements in command and control evaluation. Proc Human Factors Ergonom Soc Ann Meeting 34(17):1200–1203. https://doi.org/10.1177/154193129003401704

Mendonca D, Beroggi GE, Dv Gent et al (2006) Designing gaming simulations for the assessment of group decision support systems in emergency response. Saf Sci 44(6):523–535. https://doi.org/10.1016/j.ssci.2005.12.006

Mitschke T, Karutz H (2017) Aus-, Fort- und Weiterbildung im Bevölkerungsschutz. In: Karutz H, Geier W, Mitschke T (eds) Bevölkerungsschutz. Notfallvorsorge und Krisenmanagement in Theorie und Praxis. Springer, Berlin; Heidelberg, pp 153–166. https://doi.org/10.1007/978-3-662-44635-5_5

Murphy JP, Kurland L, Rådestad M et al (2020) Hospital incident command groups' performance during major incident simulations: a prospective observational study. Scandinavian J Trauma, Resuscitat Emerg Med 28(1):73. https://doi.org/10.1186/s13049-020-00763-4

Nilsson H, Rüter A (2008) Management of resources at major incidents and disasters in relation to patient outcome & colon; a pilot study of an educational model. Eur J Emerg Med 15(3):162–165. https://doi.org/10.1097/mej.0b013e3282f4d14b

Nilsson H, Jonson CO, Vikström T et al (2013) Simulation-assisted burn disaster planning. Burns 39(6):1122–1130. https://doi.org/10.1016/j.burns.2013.01.018

Nordhausen T, Hirt J (2020) RefHunter. Manual zur Literaturrecherche in Fachdatenbanken. Version 4.0. Tech. rep., Martin-Luther-Universität Halle-Wittenberg & FHS St. Gallen, https://refhunter.eu/files/2020/01/Manual_4.0_VFinal.pdf

NumFOCUS F (2023) Project jupyter. https://jupyter.org/

Page MJ, McKenzie JE, Bossuyt PM et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372:n71. https://doi.org/10.1136/bmj.n71

Peck M, Scullard M, Hedberg C et al (2017) Improving team performance for public health preparedness. Disaster Med Public Health Prepared 11(1):7–10. https://doi.org/10.1017/dmp.2016.65

Perry WL, Bowden FDJ (2003) Advanced metrics for network-centric naval operations. Battlespace Digitiz Netw-Centric Syst III:90–101. https://doi.org/10.1117/12.486074

Pleban RJ, Matthews MD, Salter MS et al (2002) Training and assessing complex decision-making in a virtual environment. Percept Mot Skills 94(3):871–882. https://doi.org/10.2466/pms.2002.94.3.871

Qari SH, Yusuf HR, Groseclose SL et al (2019) Public health emergency preparedness system evaluation criteria and performance metrics: a review of contributions of the CDC-funded preparedness and emergency response research centers. Disaster Med Public Health Prepared 13(3):626–638. https://doi.org/10.1017/dmp.2018.110

Reeves D, Burns J, Paris C et al (1998) Application of hand-held PC technology to support evaluation of tactical decision making performance. Proc Human Factors Ergonom Soc Ann Meeting 42(20):1422–1426. https://doi.org/10.1177/154193129804202014

Riley JM, Endsley MR, Bolstad CA et al (2006) Collaborative planning and situation awareness in Army command and control. Ergonomics 49(12–13):1139–1153. https://doi.org/10.1080/00140130600612614

Roberts AP, Stanton NA, Fay DT (2018) Go Deeper, Go Deeper: Understanding submarine command and control during the completion of dived tracking operations. Appl Ergonom 69:162–175. https://doi.org/10.1016/j.apergo.2018.02.003

Rothrock L, Cohen A, Yin J et al (2009) Analyses of team performance in a dynamic task environment. Appl Ergonom 40(4):699–706. https://doi.org/10.1016/j.apergo.2008.06.004

Roud E, Gausdal AH, Asgary A et al (2021) Outcome of collaborative emergency exercises: Differences between full-scale and tabletop exercises. J Contingencies Crisis Manag 29(2):170–184. https://doi.org/10.1111/1468-5973.12339

Rådestad M, Nilsson H, Castrén M et al (2012) Combining performance and outcome indicators can be used in a standardized way: a pilot study of two multidisciplinary, full-scale major aircraft exercises. Scandinavian J Trauma Resuscitat Emerg Med 20(1):58. https://doi.org/10.1186/1757-7241-20-58

Rönnfeldt J (2012) IT-Unterstützung in operativ-taktischen Stäben: Untersuchung der Möglichkeit existierender Programme. GRIN Verlag, Norderstedt

Rüter A, Vikstrom T (2009) Improved staff procedure skills lead to improved management skills: an observational study in an educational setting. Prehospital Disaster Med 24(5):376–379. https://doi.org/10.1017/s1049023x00007172

Rüter A, Örtenwall P, Vikström T (2007) Staff procedure skills in management groups during exercises in disaster medicine. Prehospital Disaster Med 22(4):318–321. https://doi.org/10.1017/s1049023x00004933

Salas E, Rhodenizer L, Bowers CA (2000) The design and delivery of crew resource management training: Exploiting available resources. Human Factors J Human Factors Ergonom Soc 42(3):490–511. https://doi.org/10.1518/001872000779698196

Salmon PM, Stanton NA, Ladva D, et al (2007) Measuring situation awareness during command and control activity: A comparison of measures study. Tech. rep., Human Factors Integration Defence Technology Centre, Alvington

Salmon PM, Stanton NA, Walker GH et al (2009) Measuring situation awareness in complex systems: Comparison of measures study. Int J Ind Ergon 39(3):490–500. https://doi.org/10.1016/j.ergon.2008.10.010

Savoia E, Testa MA, Biddinger PD et al (2009) Assessing public health capabilities during emergency preparedness tabletop exercises: reliability and validity of a measurement tool. Public Health Rep 124(1):138–148. https://doi.org/10.1177/003335490912400117

Savoia E, Agboola F, Biddinger PD (2012) Use of after action reports (AARs) to promote organizational and systems learning in emergency preparedness. Int J Environm Res Public Health 9(8):2949–2963. https://doi.org/10.3390/ijerph9082949

Shattuck LG, Woods DD (1997) Communication of intent in distributed supervisory control systems. Proc Human Factors Ergonom Soc Ann Meeting 41(1):259–263. https://doi.org/10.1177/107118139704100159

Silenas R, Akins R, Parrish AR et al (2008) Developing disaster preparedness competence: an experiential learning exercise for multiprofessional education. Teach Learn Med 20(1):62–68. https://doi.org/10.1080/10401330701798311

Sjöberg M, Wallenius C, Larsson G (2006) Leadership in complex, stressful rescue operations: a qualitative study. Disaster Prevention Manag Intern J 15(4):576–584. https://doi.org/10.1108/0965356061 0685901

Sjøvold E, Nissestad OA (2020) The Magellan effect - training future naval officers for mission command. Team Perform Manag Int J 26(3/4):169–182. https://doi.org/10.1108/tpm-01-2018-0008

Son C, Sasangohar F, Neville TJ et al (2020) Evaluation of work-as-done in information management of multidisciplinary incident management teams via Interaction Episode Analysis. Appl Ergonom 84:103031. https://doi.org/10.1016/j.apergo.2019.103031

Stadt Gelsenkirchen - Referat Feuerwehr (2019) ATHEBOS - Forschungsvorhaben Analyse taktisch-operativer Handlungsabweichungen (Fehler) im Einsatzgeschehen von Behörden und Organisationen mit Sicherheitsaufgaben, Teilvorhaben: Optimierung der Fehlerkultur beim stabsmäßigen Führen in der Gefahrenabwehr : Abschlussbericht. Tech. rep., Stadt Gelsenkirchen, Gelsenkirchen, https://www.tib.eu/de/suchen/id/TIBKAT%3A1700138383

Stanton NA, Harvey C (2016) Beyond human error taxonomies in assessment of risk in sociotechnical systems: a new paradigm with the EAST 'broken-links' approach. Ergonomics 60(2):1–13. https://doi.org/10.1080/00140139.2016.1232841

Stanton NA, Rothrock L, Harvey C et al (2015) Investigating information-processing performance of different command team structures in the NATO Problem Space. Ergonomics 58(12):2078–2100. https://doi.org/10.1080/00140139.2015.1046499

Sweet R (1989) Command center analyses supporting requirements specification. In: 1989 Third International Conference on Command, Control, Communications and Management Information Systems. IET, pp 27–32, https://ieeexplore.ieee.org/document/32703/

Sætrevik B, Kvamme LS (2019) Association of social network analysis with mental models in emergency management teams. OSF Preprints https://doi.org/10.31219/osf.io/puqh4

Taber MJ, McCabe J, Klein RM et al (2013) Development and evaluation of an offshore oil and gas emergency response focus board. Int J Ind Ergon 43(1):40–51. https://doi.org/10.1016/j.ergon.2012.10.006

To NT, Kato T, Imai M et al (2019) Developing a time-based evaluation method for functional exercises of emergency medical operations. Safety 5(3):49. https://doi.org/10.3390/safety5030049

Universitätsbibliothek Regensburg (2023) Datenbank-Informationssystem (DBIS). https://dbis.ur.de//index.php?bib_id=alle &colors=3 &ocolors=40 &ref=about

van Ruijven T, Mayer I, de Bruijne M (2015) Multidisciplinary coordination of on-scene command teams in virtual emergency exercises. Int J Crit Infrastruct Prot 9:13–23. https://doi.org/10.1016/j.ijcip.2015.02.005

Veenema TG, Deruggiero K, Losinski S et al (2017) Hospital administration and nursing leadership in disasters. Nursing Administ Q 41(2):151–163. https://doi.org/10.1097/naq.0000000000000224

Walker GH, Stanton NA, Salmon PM et al (2014) Human performance under two different command and control paradigms. Appl Ergonom 45(3):706–713. https://doi.org/10.1016/j.apergo.2013.09.011

Whitmore JN (2005) Assessing performance in complex team environments. Aviat Space Environ Med 76(7 Suppl):C31-3

Wilson KA, Salas E, Priest HA et al (2007) Errors in the heat of battle: taking a closer look at shared cognition breakdowns through teamwork. Human Factors J Human Factors Ergonom Soc 49(2):243–256. https://doi.org/10.1518/001872007x312478

Worm A (2001) On systems analysis and performance assessment in complex, high-risk work environments. Int J Risk Assess Manag 2(3/4):276. https://doi.org/10.1504/ijram.2001.001510

Worm A, Jenvald J, Morin M (1998) Mission efficiency analysis: evaluating and improving tactical mission performance in high-risk, time-critical operations. Saf Sci 30(1–2):79–98. https://doi.org/10.1016/s0925-7535(98)00035-6