Measures of Command & Control Performance

Nicoletta Baroutsi
PhD student
Swedish National Defense University
nicoletta.baroutsi@fhs.se

20th ICCRTS, 2015
The International Command and Control Research and Technology Symposium

Abstract

Command and Control (C2) is the function that provides direction and coordination within the complexity of crisis response situations. A number of measures have been proposed to measure the performance of C2. This is a scoping study (a work in progress) with the purpose to explore the range of the current research, to identify gaps and draw conclusions concerning the measures available. What has been discovered is: (1) the measures rely heavily on theories and the empirical support is still weak with few replication studies being made, (2) there is little known about the external validity and reliability of the reviewed measures, (3) there are no found measures that aim towards the purpose of making comparisons in unstandardized situations, which probably depends on the challenges posed by the conflicting factors in natural environments. The conclusion is: there is no method available for comparing C2 performance or measuring its progress in real life situations. If these qualities are desired, then new methods developed specifically for these real life situations need to be developed.

Introduction

During a crisis it is vital for resources, competencies and capacities to be utilized and distributed in an effective manner in order to gain control of the situation with as little loss as possible. Brehmer (2007) defined Command and Control (C2) as the function that provides direction and coordination for military operations. The definition is in this study broadened to also include civil operations, and also for direction and coordination between organizations (as opposed to only within a single organization). A variety of measures have been proposed to measure the performance of C2. This is a scoping study (a work in progress) with the purpose to explore the range of the current research, to identify gaps and draw conclusions concerning the measures available.

C2 for crisis response situations is not restricted to a single type of organization, but is instead a function that exists within a variety of organizations, and includes a wide range of variants and forms. Despite all the obvious differences there are still similarities to be found, processes and effects that are mutual, and the aim is for a measure to find and utilize these common traits. For C2 research there are two possible focus, descriptive and normative C2 research (Brehmer 2007). Descriptive research focuses on the way C2 is *actually* performed. Conversely, normative research concerns the way C2 *should* be performed. Normative research is aimed at finding critical functions that must be performed in order to achieve its purpose. Hence, if the system achieves its purpose without fulfilling a function, then it cannot be a critical function.

The processes used may differ between organizations, but the purpose of achieving direction and coordination will remain the same. This is also true for the critical functions. It is not necessary to evaluate the entire system as a whole, it is enough to evaluate the separate functions and to what extent they fulfill their separate purposes (Brehmer 2007). This implies that it is not necessary to use a measure that measures C2 as a whole, but it is also possible to use separate measures for the distinct critical functions that vital for C2.

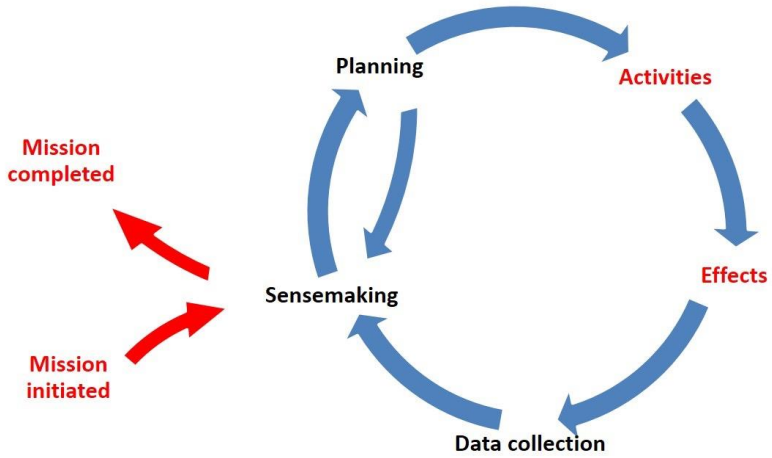

Figure 1. The DOODA-loop picturing the three critical functions of C2 and their logical relations (adaption from Brehmer, 2007).

The Dynamic OODA-loop (DOODA-loop) has identified three functions that are critical for C2. In Figure 1 is a visualization of these critical functions - *planning, sensemaking* and *data collection* - along with their logical relations. Causal and temporal relations belong at the process level and should not be confused with this visualization. The depicted relations in Figure 1 shows that the product of a certain function is a prerequisite for the next one. For example, the product of the sensemaking function is a prerequisite for the planning function. However, in reality all these functions are typically conducted simultaneously, not in a sequence. The listed functions in this model are on a high abstract level and no assumptions regarding form or process is made when it comes to how these functions are conducted in practice (Brehmer, 2007).

Purposes of measuring

When debating the applicability of performance measures it is important to be explicit about the expectations and criteria that the arguments rest upon. One important question that should be in focus while discussing measures (in all domains, not only for C2) is the intention with implementing the measure. That question being: why do we want to measure a certain concept? A measure does not have a purpose in itself, and

it is important to express these underlying intentions. For C2 performance measures it is arguable that these intentions could be:

- To measure an organization's progress over time.
- To compare the performance of one organization to another.
- To find out whether an organization fulfilled a certain degree of performance, e.g. was the performance of the organization on a satisfying level or not?

Preferably the measure should also help to distinguish where possible improvements ought to be focused. Additionally, the measure has to be possible to conduct.

Real life conditions

The outcome of a crisis depends on so many more factors than just the effort put in by the staff and volunteers. The environment where crisis management unfolds is an ever-changing entity; it is a dynamic, time-critical and high-stake situation where dynamic decision making takes place. Klein, Orasanu, Calderwood and Zsambok (1993, pp. 7-10) defined the real world by eight characteristics:

1. **Ill structured problems:** Complex causal links relate to each other, causes interact, feedback loops intertwine and so on. There is typically not one accepted procedure, and it is necessary to make a selection or invent new ways to proceed.
2. **Uncertain dynamic environments:** An incomplete world with imperfect information, some information is available while other is unavailable, ambiguous, or of poor quality.
3. **Shifting, ill-defined, or competing goals:** Well understood goals are rare outside of the laboratory setting, usually the decision maker is driven by multiple goals, some opposing each other. Typically the larger goals direct the smaller decisions.
4. **Action/feedback loops:** Series of actions stretching over time are usually needed to deal with complex problems, developing over series of events. However, the cause and effect relationship may only be loosely coupled, making it difficult to derive them back to the origin.
5. **Time pressure:** Correct decisions have to be performed, during the right time, and in the accurate order to achieve the desired result. Extensive evaluations of

multiple options are simply not feasible; only a few options are evaluated in a non-exhaustive manner before making the decision.

6. **High stakes:** The risk for severe losses affects the decision making and this becomes a methodological issue. It cannot be replicated properly in exercises or laboratories since the mere knowledge of the losses being artificial changes the mental state of the participants.

7. **Multiple players:** Many problems involve several decision makers who are actively involved. The team may include hierarchical command structures, or it may also be a flat command structure where multiple individuals act together as a single decision maker, or behave as competitors.

8. **Organizational goals and norms:** Organizations carries goals and norms that do not always coincide with the individuals' personal preferences (Klein et al., 1993, pp. 7-10).

These types of conditions are effecting the outcome of the situation and is making the task of measuring performance more complex. This will be further discussed in the next section (Limitations).

Limitations

There is a value in distinguishing between performance and effectiveness: effectiveness is related to the accomplishment of the set goals while performance is related to the team's capacities and processes (Essens, Vogelaar, Mylle, Blendell, Paris, Halpin & Baranski, 2005). This distinction is of value since the effectiveness of the team might only be circumstantial, it might just be factors in the environment that make or break the mission. For example, imagine a scenario where important supplies need to be transported from one location to another during a wartime situation. Team A initiates the mission by contacting the conflicting parties who both assures that no attacks will take place on the given route for the time of the transportation. Both the cargo and the driver arrive safely at the goal destination. In this case the team accomplished the task effectively (they reached the set goals) and showed good performance (the work processes). Team B instead sends the cargo without taking any precautionary, but the cargo still arrives safely. This team reaches high effectiveness, but the performance was not satisfactory since they did not perform any safety measures. Team C takes all the precautionary measures, but there is still a shooting and the cargo gets hit, despite the assurance from the fighting parties. In this situation the team did display good performance since they correctly executed the working

processes, even if the effectiveness fell through. In this study it is the performance that is of interest, not the effectiveness.

For the current review, the interest lies within ways of measuring the actual performance of an organization during either a drill or a sharp situation. This excludes all measures related to risk assessments and other prerequisites (e.g. group qualities) that can be related to performance, since it is not possible to beforehand measure actual performance but only possibilities. These limitations also prevent the study to get into the territory of virtual agents in simulated environments, since the applicability of these measures in real life is questionable.

Scoping studies

This paper is based on the scoping study methodology, a methodology that aims toward summarizing a wide array of literature. It is suited when the purpose of a study e.g. is (Arksey & O'Malley, 2005):

- To explore the range, nature and extent of the research, without describing the findings in detail.
- To identify gaps and draw conclusions concerning the overall state in the chosen research area.

However, quality assessment is not part of the scoping study so this choice of study will not necessarily identify gaps that depend on poor quality.

Summary

The purpose of the study is to explore the range of the measures available proposed to measure C2 performance, to identify gaps and draw conclusions concerning these measures. C2 has been defined as the function that provides direction and coordination for military operations (Brehmer, 2007). This includes both civil and military operations, as well as within and between organizations.

A measure needs an intention; it has to fulfill a purpose for the user. Three possible intentions for a C2 measure have been presented: (1) to measure progress over time or (2) to compare one organization to another (3) to find out whether a certain degree of performance have been fulfilled. When measuring C2 it is both possible to measure C2 as a whole, or to measure the critical functions independently from one another.

A distinction has been made between effectiveness and performance, with performance being the focus of this review. Also, the measure has to be applicable in either a sharp situation or a drill, and all types of prerequisites are excluded.

Method

This paper explains a scoping study (work in progress) that systematically searches through the literature. This work will be complemented with additional searches in databases, hand searching journals, and interviewing relevant scientists. At present, the search has been conducted in the Scopus database and complemented through snowballing, i.e. looking through the references of the relevant papers, and using tips from scientists in related domains.

When initiating the scoping study the purpose was to investigate the measures that are applicable when it comes to evaluating C2 performance. The limitations were not yet established, partially because of a lack of knowledge concerning the sheer amount of work available in the selected domain. If few papers were to be found, then a wider search range would be selected, but if relevant papers would be easy to find, then a more narrow scope would be a better choice.

The methodology builds upon the description put forth by Arksey & O'Malley (2005), which includes following steps:

- Identifying the research question
- Identifying relevant studies
- Study selection
- Charting the data
- Collating, summarizing and reporting the results

The steps were conducted in an iterative fashion and each step was revisited when needed. A summary of the steps taken can be seen in Figure 2, and an elaborated description of each step will follow.

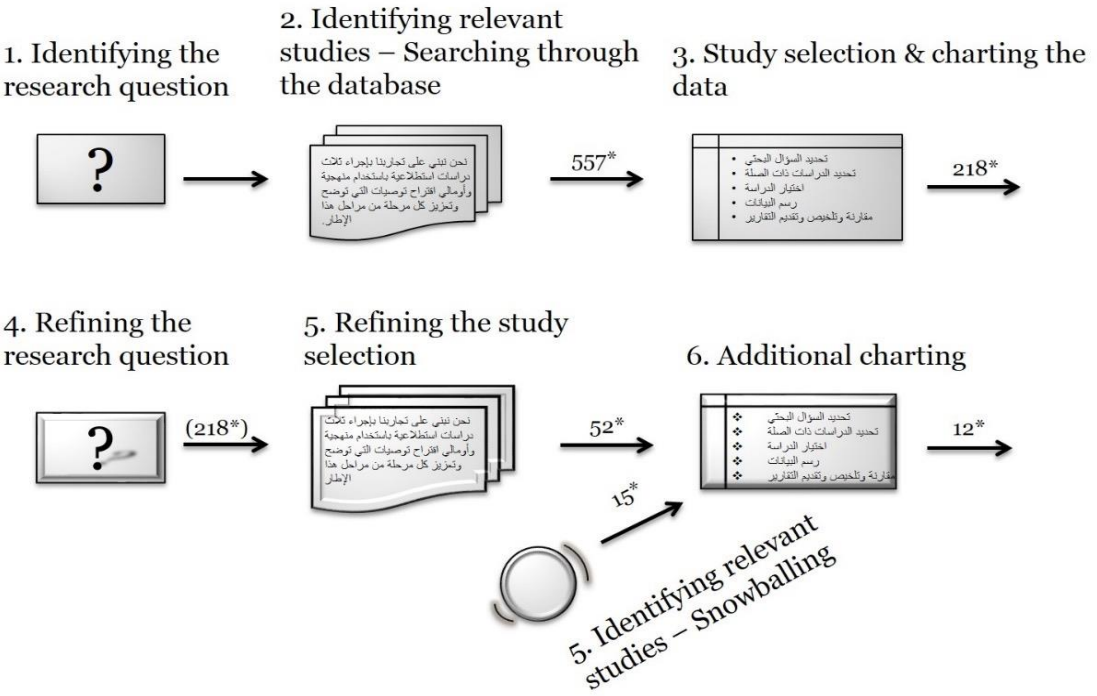

*Figure 2. Summary of the steps conducted in this work in progress paper. * Number of articles transferring into the next step.*

Step 1: Identifying the research question

The research question was complemented with two sub-questions designed to direct the choice of what literature to include in the study. Any literature that might prove helpful at answering any of the following questions is included in the initial selection. However, these limitations were altered during the course of the work, see section 'Limitations' for final version.

How can C2 performance in crisis response situations be measured?

- What are the needed pre-requisites for effective C2?
- What indicators have been found that relate to effective C2?

Step 2: Identifying the relevant studies – Searching through the database

Relevant keywords were identified within the research question, namely 'command & control', 'performance' and 'crisis'. These keywords were then used as a starting point for the upcoming searches, and complemented with synonyms. The complete set of keywords can be seen in Table 1.

Table 1. Complete set of keywords used in the Scoping Study, here complemented with a grade reflecting their usefulness. ***Grade A:*** *Very interesting results.* ***Grade B:*** *Interesting results, but few hits.* ***Grade C:*** *Interesting results.* ***Grade D:*** *Too few hits, but usable if keyword 2 is removed.* ***Grade E:*** *Borderline (skim through).* ***Grade F:*** *Not interesting.* ***Grade G:*** *Too many hits, search is not manageable.*

Keyword 1	Grade	Keyword 2	Grade	Keyword 3	Grade
Command and control	**A**	Performance	**A**	Crisis	**A**
Collaboration	**B**	Analysis	**A**	Disaster	**A**
Response management	**B**	Evaluation	**C**	Multi organisational	**D**
Coordination	**C**	Assessment	**C**	Catastrophe	**D**
Cooperation	**F**	Measur*	**C**	Accident	**E**
Management	**G**	Effect* (iveness)	**C**	Emergency	**E**
C2	**G**	Capabilit*	**C**		
		Audit	**F**		

The searches were performed in Scopus by combining the words in the three keyword categories, i.e. one word from each category. The titles in the selected searches were read through and relevant hits marked. All search queries were listed in a separate document together with relevant information (keyword, date of the performed searches, search query, number of titles, number of relevant titles, perceived usefulness, and comments). When completed a total of 557 abstracts were downloaded.

Step 3: Study selection and charting the data

All the abstracts were read through and graded (low, medium and high relevance). Highly relevant articles were downloaded and saved in Mendeley, a citation management software. These papers were read through and searched for specified content:

1. Category (risk, measure, evaluations, qualitative etc.)
2. Purpose
3. Empirical support
4. Metric (expert evaluations, social networks, indicator etc.)
5. Concept (collaboration, C2, decision making etc.)
6. Statistical method
7. Domain (medicine, response etc.)
8. Results
9. Comments
10. Relevance (not relevant, questionable, relevant, highly relevant)
11. Concerns

This charting narrowed down the articles to 218 hits.

Step 4: Refining the research question

After the initial charting of the literature a choice was made to narrow the scope of the study. The material outside of the new limitations was assessed to be non-contributable to the initial purpose - to investigate measures applicable for evaluating C2 performance. These are the final limitations, that are also presented under the 'Limitations'.

How can C2 performance in crisis response situations be measured?

- **For the current purpose the focus is on performance, not effectiveness**: Effectiveness is related to the accomplishment of set goals, and includes many variables not related to the team. Performance is instead related to the team's capacities and processes (Essens, Vogelaar, Mylle, Blendell, Paris, Halpin & Baranski, 2005).
- **All measures related to risk assessments and other prerequisites are excluded** (e.g. group qualities), even if they can be related to performance.
- **The measure should be applicable during either a drill or sharp situations.** These limitations prevent the study to get into the territory of virtual agents in simulated environments.

Step 5: Refining the study selection / Identifying relevant studies - Snowballing

The study selection was narrowed based on the refined limitations. This resulted in a total of 52 papers that were complemented with 15 new papers found through snowballing, hence a total of 67 papers.

Step 6: Additional charting

Additional relevant information was identified. The sought out information included:

1. Validation: How the measure had been validated
2. Logical steps leading from concept to performance: How the relation between the concept and performance is described.
3. Analysis: How the data was analyzed
4. Data collection: How the data was collected
5. Measure/metrics: How the concept is measured and what type of metrics that was used.
6. Domain: In what domain the measure had been applied

7. Comments
8. Level of relevance: High, medium or low

Only 12 papers were found to be suited for the current purpose. No time-span was limiting the searches, still the relevant articles only dated back to 2002. The work will progress by conducting additional searches in Scopus, hand searching specific journals and interviewing relevant scientists.

Results

The included measures relate to C2 in different ways: Indicators and Measures of Effectiveness (MoE) relate to C2 within organizations, social network relates to C2 between organizations, and awareness relates to the sensemaking function in the DOODA-loop. The choice to include awareness will be elaborated under 'Awareness'. This wide definition of C2 was chosen since the purpose of the study is to explore the range of the measures available, hence an inclusive approach was chosen. The included papers are presented in Table 2.

Table 2. These are the titles and authors of the selected papers.

Nr	Authors	Title	Measure
1	Djalali, Carenzo, Ragazzoni, Azzaretto, Petrino, Della Corte & Ingrassia	Does Hospital Disaster Preparedness Predict Response Performance During a Full-scale Exercise? A Pilot Study	Indicators
2	Abbasi, Hossain, Hamra, & Owen	Social networks perspective of firefighters' adaptive behaviour and coordination among them	Social network
3	Gryth, Radestad, Nilsson, Nerf, Svensson, Castren, & Rüter	Evaluation of Medical Command and Control Using Performance Indicators in a Full-Scale , Major Aircraft Accident Exercise	Indicators
4	Hossain & Kit Guan	Modelling coordination in hospital emergency departments through social network analysis	Social network
5	Höglund & Berggren	Using shared priorities to measure shared situation awareness	Awareness
6	Kapucu, Augustin, & Garayev	Interstate Partnerships in Emergency Management: Emergency Management Assistance Compact in Response to Catastrophic Disasters	Social network

7	Macmillan, Paley, Entin & Entin	Questionnaires for Distributed Assessment of Team Mutual Awareness. In 'Handbook of Human Factors Methods'	Awareness
8	McGuinness & Foy	A subjective measure of SA: the Crew Awareness Rating Scale (CARS).	Awareness
9	NATO CCRP	Code of best practice for command and control assessment	MoE
10	Nilsson & Rüter	Management of resources at major incidents and disasters in relation to patient outcome : a pilot study of an educational model	Indicators
11	Rådestad, Nilsson, Castrén, Svensson, Rüter, & Gryth	Combining performance and outcome indicators can be used in a standardized way: a pilot study of two multidisciplinary, full-scale major aircraft exercises	Indicators
12	Sproles	Formulating measures of effectiveness	MoE

Indicators

The most widely used C2 measure is indicators (Djalali et al. , Gryth et al., 2010, 2014, Nilsson & Rüter, 2008, Rådestad et al. 2012). Usually the method includes domain experts observing an exercise for predefined indicators while also assigning the indicator with ordinal values that reflected the state of a specific behavior. For example, 0 – not timely or adequate, 1 – somewhat correct, and 2 – timely and adequate. These indicators are also often combined with the time frame in which it is executed. For example, minutes from alert (Gryth et al. 2010). The vast majority of the papers using these techniques are related to the medical field adopting the technique in different areas, ranging from prehospital C2 to C2 within the hospital environment. One of the five articles applied templates of previously used indicators (Gryth et al. 2010), while the rest created them based on the current situation. You could ask why this favourising of the medical field has emerged, but the more suited question would be whether this technique could be adopted to other domains as well. It might however not be easily transferable; the indicators are based on standardized, accepted methods and doctrines, and the staff is expected to respond according to these predefined procedures.

Awareness

The choice to include awareness might not be clear, but the decision was based on multiple reasons. First, the Shared Priorities measure (Höglund & Berggren) uses 'command and control' as a keyword and therefor appeared among the searches (the other two articles were found through snowballing). Also, the theoretical framework (the DOODA-loop) specifies that the critical functions can also be measured separately if one wants to measure C2. Awareness would here be related to the sensemaking function. These reasons, in combination with the choice of methodology (scoping

study) that is supposed to be about range and breadth, supported the choice to include this category.

Shared situation awareness and mutual awareness are similar concepts that relate to how a team or a group of people apprehends each other and their surroundings. The team score is calculated by aggregating or comparing results of the individuals. A common approach is to use self-assessment questionnaires asking the participants about their own and their teammate's workload, as well as process or performance related questions. These approaches can be seen for CARS (McGuinness & Foy, 2000) and DATMA (MacMillan et al. 2005, p. 51-71). Additions to these Likert-scale questionnaires include recollection and questions regarding a salient event, a technique applied in DATMA. The CARS and DATMA measures have been found to co-vary with performance, but a recent study suggests that they might be stronger related to task complexity than team performance (Baroutsi, Berggren, Nählinder & Johansson, 2013). In this study a computer simulation presented trained teams and untrained teams with scenarios of varying difficulty. The used self-assessment techniques, CARS and DATMA, could however only differentiate between variations in the scenarios but not between trained and untrained teams. The same study found that another measure, Shared Priorities (Höglund & Berggren), instead was able to differentiate between the types of teams but not the types of scenarios. For this measure the participants are all asked to individually write down and rank five goals that the team is attempting to achieve at the moment. These lists are collected and scrambled, and the members get to rank each other's sets of goals. The sets of ranked lists are then compared within the team. Shared Priorities rests on the belief that teams who prioritize tasks similarly, and share the same situation awareness, will perform better. Shared Priorities still needs further validation and it also comes with scaling issues, it is today not applicable for larger groups.

Measures of Effectiveness

MoEs are used within the military domain to evaluate whether a certain goal or sub goal has been achieved. It is a way of determining whether a satisfactory solution has been obtained (Sproles, 2002). The earliest references of MoEs can be found in 1985, consisting of a series of workshops developed by the U.S. Naval Postgraduate School. The furthest elaborated version seems to be brought out by NATO that has developed a series of related measures that strive to quantify effectiveness on different levels within what they call Measures of Merits (NATO CCRP, 2002). These include:

- Measures of Policy Effectiveness: societal and policy outcomes
- Measures of Force Effectiveness: to what degree a force meets its objectives

- Measures of C2 Effectiveness: the impact of C2 systems within a operational context
- Measures of Performance: relates to a systems characteristics, structure and behavior
- Dimensional Parameters: inherent characteristics of the physical C2 system

This measure has been criticized for the confusion on how to properly formulate MoEs and that they are relying on the creativity of the user. There are no good management techniques but creation is instead based on common sense, MoEs should however be able to be tested in some way (Sproles, 2002). Sproles (2002) presents some examples of MoEs: the number of merchant ships saved, amount of cargo saved, and time to establish a temporary base. Attempts to formulate the proper steps have been made, but they have yet to be accepted by the practitioners.

Social networks

Social networks are used to examine attributes of the relations between organizations or individuals. The organizations/individuals are represented by *nodes* and the relationships within the network are represented by *ties.* Relationships are calculated through recorded communication, e.g. emails, phone calls etc. Many metrics can be derived through the use of these networks and each reviewed paper (Abbasi et al., 2010, Kapucu et al. 2009, Hossain & Kit Guan, 2010) has chosen a unique set, including metrics of density, degree centrality, betweenness centrality, individual tie strength, teams' internal and external tie strength and so on. While there have been studies relating patterns of coordination to high performance only one attempts to measure the effectiveness of an organization through the social network, and also compares one sharp situation to another based on these results (Hossain & Kit Guan, 2010). This article compares the coordination effectiveness during two different hurricanes, but as they themselves stated, this difference is affected by the different sizes of the catastrophes.

Analysis

Before entering the analysis, a reminder that this is a work in progress and that the analysis is yet only based on 12 papers.

Initially three purposes for measuring C2 were presented, and the measures found were mainly aimed at accomplishing the third purpose; to find out whether an organization fulfilled a certain degree of performance. Included here were some of the papers that applied indicators as well as the MoEs. However, the research has not yet been able to establish benchmarks that have empirical support, but is rather supported

by the creativity, skill and 'common sense' of the individual. Some discussions have been presented on weighting different indicators in order to assign a higher value to those that are more critical (Gryth et al., 2010). How this would actually be accomplished has not been clarified. The medical domain seems to be further ahead, based on the number of articles related to this domain as well as the presented discussions regarding future endeavors wished to accomplish. Unfortunately, the method developed for this field is not directly transferable since they are backed up by standardized and accepted doctrines. Therefore one would first need to develop these doctrines if this would be the chosen way forward. So far, no presentation of applied MoEs during drills or sharp situations have been found.

Other measures, including some of the indicators and measures aimed at measuring situation awareness, did not offer such benchmarks but rather used the measures to make comparisons. These comparisons are dependent on the stability of the environment since they are influenced by all the distracting factors found in the real world. Hence, they are well suited for standardized training and simulations, but they cannot be applied during sharp situations or on unstandardized scenarios. Among the papers only one of them made comparisons of real life situations (Hossain & Kit Guan, 2010). Unfortunately the comparison is hampered since the size of both the organizations and the magnitude of the crises differed, this dilemma is brought forward in the paper as well.

The validity of the measures today relies heavily on theoretical support; including theories concerning mediating concepts that relate to performance, or accepted doctrines. For the studies that applied indicators and MoEs (also social networks) used unique sets of indicators or metrics to fulfill the stated purposes. This raises questions to whether they can be applied in multiple settings or scenarios. Also, replication studies seem to be rare. Hence, the external validity and the reliability of the measures are still unclear.

The medical and military domains are overrepresented among the studies and the majority of the attempts use indicators or MoEs to evaluate C2 performance.

Suggestions and conclusions

Today the measures are relying on a theoretical underpinning, but the external validity and reliability of the measures are still uncertain. To overcome these problems the measures would benefit from the discussion being brought back to these fundamentals that the measures rest upon, including replication, external validity, reliability and so on. The specified domain is narrow and still quite young, but the problems facing this

domain are not unique and they should not be ignored. Measures have a value in that they create a ground to make comparisons, but these qualities are today only present for stabilized environments. This dilemma is not unique to C2 measures, it would be fairer to say that this is the core dilemma when trying to measure human processes and effects in these dynamic, time-critical, high stake environments that is real life. The same problem would emerge also when trying to follow the development of an organization over time, since the result might only reflect differences in the complexity of the different situations.

How can performance in dynamic, time-critical, high stake situations be made comparable? All the influencing factors make comparisons problematic. The straight-forward answer is to standardize exercises, i.e. to keep the conflicting factors constant, which would allow for the subjects' progress to be measured. Any progress could however be affected by task familiarity, and this method would also not be applicable in sharp situations. Methods for measuring C2 stem from the same principles as when studies were conducted in laboratories, but these techniques cannot be transferred into the complexity of the real world. If these are qualities that are desired of a measure, then new methods derived specifically for these dynamic, real life situations needs to be developed. Standardizing scenarios is beneficial for developing measures, but this would limit their usefulness.

References

Abbasi, A., Hossain, L., Hamra, J., & Owen, C. (2010). Social networks perspective of firefighters' adaptive behaviour and coordination among them. *IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing*, 819–824.

Arksey, H. & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal Social Research Methodology, 8 (1),* 19-32.

Baroutsi, N., Berggren, P., Nählinder, S. & Johansson, B. (2013). *Training teams to collaborate as cohesive units.* (FOI-R--3830--SE). Linköping.

Brehmer, B. (2007). Understanding the functions of C2 is the key to progress. *International C2 Journal*, 1, 211-232.

Djalali, A., Carenzo, L., Ragazzoni, L., Azzaretto, M., Petrino, R., Della Corte, F., & Ingrassia, P. L. (2014). Does hospital disaster preparedness predict response performance during a full-scale exercise? A pilot study. *Prehospital and Disaster Medicine*, *29*(5), 441–447.

Gryth, D., Radestad, M., Nilsson, H., Nerf, O., Svensson, L., Castren, M., & Rüter, A. (2010). Evaluation of Medical Command and Control Using Performance Indicators in a Full-Scale, Major Aircraft Accident Exercise. *Prehospital and Disaster Medicine*, *25*(2), 118–125.

Höglund, F., & Berggren, P. (2010). Using shared priorities to measure shared situation awareness. In *Proceedings of the 7th International ISCRAM conference*.

Hollnagel, E. (2012) *FRAM, the functional resonance analysis method: Modelling complex socio-technical systems.* Farnham, Surrey, UK England: Ashgate.

Hossain, L., & Kit Guan, D. C. (2012). Modelling coordination in hospital emergency departments through social network analysis. *Disasters*, *36*(2), 338–365.

Kapucu, N., Augustin, M. E., & Garayev, V. (2009). Interstate partnerships in emergency management: Emergency management assistance compact in response to catastrophic disasters. *Public Administration Review*, *69*(2), 297–314.

Klein, G., Orasanu, J., Calderwood, R., & Zsambok, C. (1993). *Decision making in action.* Norwood, New Jersey: Ablex Publishing Corporation.

Macmillan, J., Paley, M. J., Entin, E. B., & Entin, E. E. (2005). Questionnaires for Distributed Assessment of Team Mutual Awareness. In N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendricks. *Handbook of Human Factors Methods*. London: Taylor and Francis.

Manser, T., Harrison, T. K., Gaba, D. M., & Howard, S. K. (2009). Coordination patterns related to high clinical performance in a simulated anesthetic crisis. *Anesthesia and Analgesia*, *108*(5), 1606–15.

McGuinness, B., & Foy, L. (2000). *A subjective measure of SA: the Crew Awareness Rating Scale (CARS)*. In The Human Performance, Situational Awareness an Automation Conference.

Meirina, C., Levchuk, G. M., Ruan, S., Pattipati, K. R., & Popp, R. L. (2006). Normative framework and computational models for simulating and assessing command and control processes. *Simulation Modelling Practice and Theory*, *14*(4), 454–479.

Nilsson, H., & Rüter, A. (2008). *Management of resources at major incidents and disasters in relation to patient outcome : a pilot study of an educational model*.

Rådestad, M., Nilsson, H., Castrén, M., Svensson, L., Rüter, A., & Gryth, D. (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 Journal of Trauma, Resuscitation and Emergency Medicine*, *20*(58).

Sproles, N. (2002). Formulating measures of effectiveness. *Systems Engineering*, *5*(4), 253–263.

NATO CCRP (2002). *Code of best practice for command and control assessment*.

Essens, P., Vogelaar A., Mylle, J., Blendell, C., Paris, C., Halpin, S., & Baranski, J. (2005) *Military command team effectiveness: Model and instrument for assessment and improvement*. ISBN 90-6464-266-4.