14th ICCRTS

"C2 and Agility"

Evidence-Based C2 Metrics: A Survey

Topic 8: C2 Assessment Tools and Metrics

Anthony H. Dekker

Defence Science and Technology Organisation (DSTO)
Canberra, Australia
dekker@acm.org

Paper #193

Evidence-Based C2 Metrics: A Survey

Anthony H. Dekker
Defence Science and Technology Organisation (DSTO)
Canberra, Australia
dekker@acm.org

Abstract

In this article, we survey some of the recent literature on C2 Metrics, and list a number of metrics for which there is published literature supporting their validity and feasibility. The metrics are classified by whether they measure process properties, document properties, or human cognitive properties.

1. Introduction

In today's complex military operations, it is important that headquarters perform as effectively as possible. This requires learning lessons from exercises and past operations. Using Command and Control (C2) metrics to assess headquarters performance is an important part of such organisational learning.

It might seem that that we can simply measure the success of an operation or exercise, and use that to assess headquarters performance. However, this introduces confounding factors – such as equipment problems and unanticipated events – which obscure the success or failure of the headquarters itself. In command-post exercises, the plan produced by the headquarters may never be executed, and so this measure becomes simply impossible. For both reasons, it is important to have C2 metrics that can be used to assess the performance of a headquarters in isolation from the forces it is commanding.

C2 metrics used to assess headquarters must be both *feasible* and *valid*. Some excellent candidate metrics are simply too difficult to apply – they are not feasible. For example, the quality of the plan produced by a headquarters can in theory be assessed in the same way that judging is applied in some Olympic sports: assemble a panel of 5 to 7 expert judges, have them each rate quality on a 0 to 10 scale, discard the highest and lowest scores, and average the rest. In practice, assembling a suitably qualified judging panel is extremely difficult, particularly at the joint level, since only experienced former commanders are qualified to act as judges. Other obstacles to feasibility include excessive analyst workload or interruption to headquarters operation.

Some other metrics are easy to apply, but useless, because they are not valid. To take an absurd example, measuring the length of officers' haircuts is simple, but bears no relationship to headquarters performance. Indeed, to use such an invalid metric is actively harmful, since it encourages staff to divert energy into "looking good" from the point of view of that metric, with a resulting shift in focus away from the task at hand.

In this brief survey paper, we will summarise some of the recent literature on C2 metrics, concentrating on metrics where there is published evidence supporting validity and feasibility. At the least, the metrics should have been successfully applied in practice, and techniques for applying the metrics should have been described in published reports. Our goal in doing this is to identify the international "state of the art" in this field, as a starting point for moving forward.

A headquarters can be described as a group of *people* who go through a *process* to produce *documents* (such as plans, orders, briefing presentations, and messages). We can therefore divide metrics into three groups: *cognitive metrics*, which assess the understanding of the people involved; *process metrics*, which assess the efficiency of the process; and *document metrics*, which assess the outputs of the process.

2. Document Metrics

One of the main functions of military headquarters is to produce various kinds of *documents*, including plans, orders, briefing presentations, and messages. The 1940 order from General Heinz Guderian shown in Figure 1 provides an example. Consequently, one way of examining the performance of a headquarters is to use metrics which assess the quality of these documents. This has the advantage is that it can be done after an exercise is complete, or outside the active area of the headquarters, and therefore need not disturb the headquarters in any way. Document metrics are often analogous to time-honoured techniques from academia, which lends support to their validity. We list several such metrics below, together with simple estimates of their validity and feasibility. These estimates are subjectively derived (using a 0 to 3 star rating scale), but based on the evidence provided in publications.

Understandability – Restatement (validity ★★ feasibility ★★)

One of the most important quality measures for documents produced by a headquarters is *understandability*. Do the intended users of these documents – typically commanders and staff in subordinate headquarters – understand what they say? If they do not, they will not be able to carry out the intent of those documents effectively. An effective, though somewhat time-consuming, way to measure understandability is to ask all users of the document to articulate in their own words the key points of the document, and to compare this against a list compiled with the assistance of the author (this is analogous to what students are often asked in an ordinary examination). The Republic of Singapore Armed Forces have successfully used this method for assessing subordinate's understanding of commander's intent in Brigade and Divisional Command Post exercises (Cheah and Fong 2006).[1]

Understandability – True/False Questions (validity ★★★ feasibility ★★)

An alternative approach is to prepare a list of true/false questions or "probes" related to the document at hand, such as "Sharing information with NGOs (Non-Government Organisations) is in line with the Commander's intent to avoid civilian casualties [T/F]." This method has been used in a US/Singapore Coalition Task Force exercise, where it was able to detect statistically significant differences between approaches (Thomas *et al*. 2007). It has the advantage of being less subjective than the restatement method, although analyst effort is still required in assembling an appropriate scenario-specific questionnaire.

The option of asking users of a document to self-assess their level of understanding is unlikely to be valid. People in general cannot accurately assess their own skill levels (Kruger and Dunning 1999), and objective measures of understanding are to be preferred.

---

[1] The Republic of Singapore Armed Forces have explored C2 metrics extensively, in collaboration with several other countries, and also actively published this work in forums such as ICCRTS.

Data Sources – Number (validity ★★ feasibility ★★)

A method often used in academic publishing (at least informally) is to assess a document by the number of data sources used in its production. A greater number of data sources reflects a better-informed author or authors, and presumably a higher-quality document. Counting data sources is easiest if they are explicitly cited, as in academic publications. However, data sources can also be inferred by scanning a document, although this is more time-consuming. The order by Guderian in Figure 1 used five data sources, for example (reports from two subordinate divisions, reports from XIV Corps, reports from higher headquarters, and ISR reports).

Data Sources – Timeliness (validity ★★ feasibility ★★)

It is also possible to assess a document by the recency of its data sources. In academic publications, for example, the year of cited publications indicates recency. In documents such as orders, an implicit or explicit date or time would indicate recency, such as "yesterday rain was falling over the target area" or "at 08:00 this morning rain was falling over the target area." For the order in Figure 1, all data sources seem to have been from earlier the same day. The list of dates or times can be easily converted to a single metric by taking the median value. The median date of references for the present work is 2005, for example.

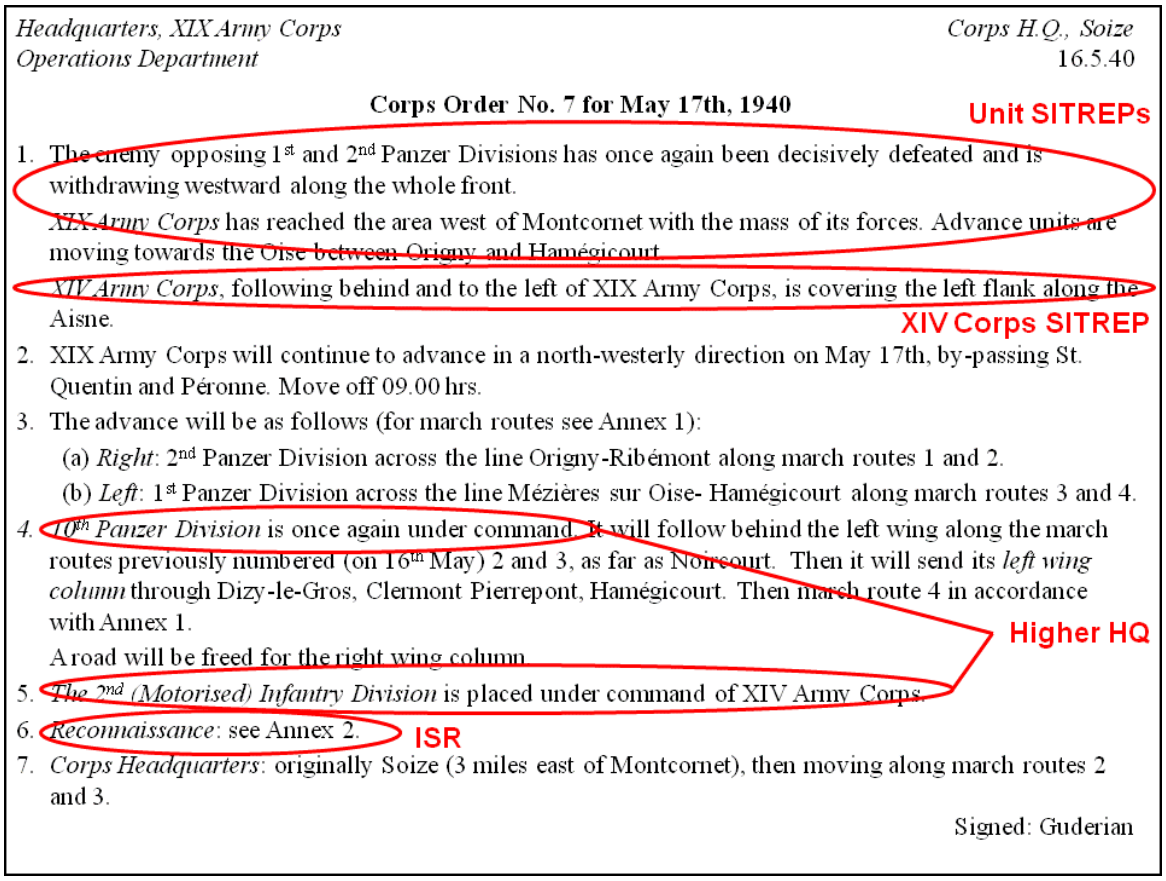

Figure 1: Example corps order produced by General Heinz Guderian in 1940, issued on the evening of 16 May, based on events of the day (from Appendix IX to Heinz Guderian's 1952 autobiography *Panzer Leader*). The five data sources used are highlighted.

3. Process Metrics

The second main approach to C2 metrics is to enter a headquarters and measure aspects of the process going on within it. This can be done by a human observer, which runs the risk of interfering with operations (although many military personnel are accustomed to such observation). Alternatively, electronic systems such as chat rooms and e-mail can be monitored without interfering with headquarters operations.

Process – Timing (validity ★★★ feasibility ★★★)

The easiest aspect of headquarters operation to assess is *time*. This includes measuring the time taken to react to events, the time taken to perform tasks (including the time taken to produce the final plan), and the throughput of tasks per hour (NATO Research and Technology Organisation 2002, p 102). The only thing required here is to record the time of significant events, and of the start and end of key processes. Given the importance of tempo in military operations, more rapid reaction and faster throughput will, all other things being equal, have a positive impact on performance. However, because of the danger of fast and sloppy work, time factors must always be measured in conjunction with quality factors.

Process – Breadth (validity ★ feasibility ★★★)

Another easy form of assessment is to count the number of courses of action (or contingencies, or potential consequences) considered in planning. Such metrics are recommended in the US *Joint C2 Functional Concept* (US Department of Defense 2004, Appendix C). However, simply counting courses of action considered does not distinguish between those worked out in detail, and those merely sketched in order to satisfy the requirement for a certain number of courses of action. Consequently, the validity of simply counting courses of action is not clear.

Process – Workload (validity ★★ feasibility ★★★)

The Republic of Singapore Armed Forces (Cheah and Fong 2006) have measured workload in headquarters using the NASA Task Load Index (Hart 2006). This is a widely used subjective measure of workload, based on a weighted average of ratings on six subscales (Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort, and Frustration). Because of the extensive literature on this metric, it can be used to identify staff who are overloaded.

Process – Teamwork (validity ★★ feasibility ★★★)

Failure of teamwork among military personnel has a negative effect on planning (Dekker 2006), and the quality of teamwork is therefore very much worth assessing. The Republic of Singapore Armed Forces (Cheah and Fong 2006) has used the "Star Performing Teams" questionnaire from NZ-based firm Team Management Services (Team Management Services 2008). This questionnaire provides a subjective measure of how well team members think their team is performing, using ten categories of questions, such as "The team is aligned internally in terms of having a common mindset [1/2/3/4/5/6]." Although assessing teamwork is very important, this particular questionnaire has not been widely used in the C2 area, given the range of other options available (Kyne *et al*. 2002). However, it is not clear which of these teamwork measures is best.

Process – Interoperability (validity ★★ feasibility ★★★)

An important aspect of planning at the joint level is *organisational interoperability*. Military headquarters must work with multiple nations, multiple services, and multiple agencies. Ford *et al*. (2007) survey a number of metrics for interoperability, of which the Australian Organisational Interoperability Maturity (OIM) model is highlighted as particularly well-cited. The OIM (Clark and Moon 2001) assesses organisational interoperability between two agencies using four categories (Preparedness, Understanding, Command Style, and Ethos) and five levels (Independent, Cooperative, Collaborative, Combined, and Unified). The overall interoperability level for a pair of agencies is the lowest among the four categories.

Process – Aggregated Measures (validity ★ feasibility ★★)

There are a number of long-standing aggregate measures for C2, including the Headquarters Effectiveness Assessment System (HEAT) and the Army Command and Control Evaluation System (ACCES). The difficulty with these systems is that the meaning of the final score is unclear, and the *NATO Code of Best Practice for C2 Assessment* (NATO Research and Technology Organisation 2002) suggests that such measures "have limitations in the diagnosis of C2 success or failure."

Organisational Network – Average Degree (validity ★ feasibility ★★★)

The techniques of Organisational Network Analysis or Social Network Analysis (Wasserman and Faust 1994) can shed light on the operation of an organisation such as a headquarters (Desouza *et al*. 2008). Organisational Network Analysis involves tracking communication between staff. This can be done by monitoring e-mail traffic, downloading logs from a telephone PABX, distributing questionnaires asking who people communicate with, or arranging for all people who communicate within a given time window to exchange business cards. Once these "links" between people have been recorded, they can be summarised in a diagram, such as Figure 2.

Several different metrics can be calculated from the network of organisational communication. Jarvis (2005) reports such an analysis of e-mail traffic for a US/UK naval exercise. The simplest metric that can be inferred from the organisational network is the average number of contacts per person (the *average degree*). For the network in Figure 2, this is 15.2 (with a range of 3 to 37 contacts per person). However, the validity of this metric is not entirely clear, since communication with many people can represent either effective information-gathering, or ineffective time-wasting.

Organisational Network – Average Distance (validity ★★★ feasibility ★★★)

A better metric to derive from Organisational Network Analysis is the *average distance* between people in the network. The average distance is computed by measuring the number of links in the shortest path between all pairs of people, and averaging these numbers (for Figure 2, this gives 1.80). Several studies, such as Dekker (2007), show that a low average distance is associated with more effective and more rapid transfer of information within an organisation.

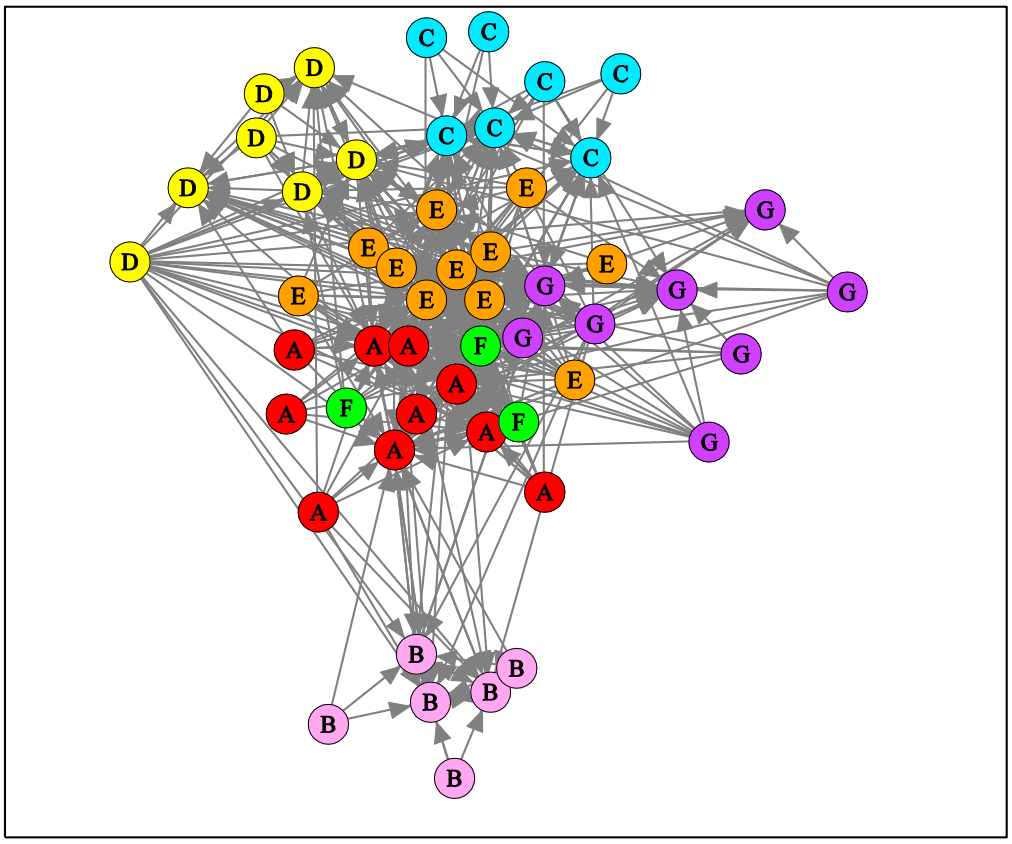

Figure 2: Organisational Network diagram for a military organisation, with branches colour-coded. A number of metrics can be calculated from such a network.

4. Cognitive Metrics

The third kind of C2 metrics are *cognitive metrics*. Unlike document metrics (which assess the outputs of the headquarters process), and process metrics (which assess the process itself), cognitive metrics assess what is going on in the minds of the staff. The understandability metrics and workload metrics described above can therefore also be considered to be cognitive metrics. Cognitive metrics are the most difficult to apply, and the most debated. In this section, we will concentrate on metrics for situational awareness (SA), the area in which there is most agreement.

SA – SAGAT (validity ★★ feasibility ★)

A widely used measure of situational awareness for individual people is the SA Global Assessment Technique (SAGAT), which goes back to 1995. The Republic of Singapore Armed Forces have used this in exercises (Cheah *et al*. 2005), as have many other organisations (Natter *et al*. 2008, Stanners and French 2005). It is an objective measure, asking participants situation-specific questions about their perceptions of what is going on, and assessing these against ground truth. Studies have shown that the SAGAT correlates with performance, which more subjective measures of SA do not (Natter *et al*. 2008). However, a significant disadvantage of the SAGAT approach is that it requires regular "freezes" or interruptions of activity, which are not practical in a large headquarters. It has also been subject to more general criticism (Nofi 2000).

SA – True/False Questions (validity ★★★ feasibility ★★)

A better method for assessing SA is to use true/false "probe" questions about the scenario, much like those discussed under document understandability above. For SA issues at the operational (rather than tactical) level, true/false probes provide greater objectivity than more open-ended questions. The QUantitative Analysis of Situation Awareness (QUASA™) method is one way of administering such probes, and has been used successfully in a US/Singapore Coalition Task Force exercise (Thomas *et al*. 2007).

SA – Team SA (validity ★★ feasibility ★★)

Even more difficult than individual SA is the problem of assessing SA in groups or teams, and there is no consensus on how this should be done (Natter *et al*. 2008). However, the use of true/false probe questions extends naturally to a team environment, if for each question the *worst answer in the team* is used. The alternative of using agreement between team members – as in Perla *et al*. (2000) – is less desirable, because teams can agree while still being wrong (Dekker 2006).

5. Conclusions

A headquarters can be described as a group of *people* who go through a *process* to produce *documents* (such as plans, orders, briefing presentations, and messages). There are therefore three kinds of C2 metrics: *document metrics* (which assess the outputs of the process), *process metrics* (which assess the efficiency of the process), and *cognitive metrics* (which assess the understanding of the people involved). We have briefly looked at 15 example metrics reported in recent published literature. Table 1 summarises these metrics.

It is important that C2 metrics used to assess headquarters be both *feasible* and *valid*. They should involve neither too much analyst workload nor too much disruption of headquarters activities. Their validity should also be supported by published studies. Table 1 gives estimates for the validity and feasibility of the 15 metrics examined here.

The most promising metrics among those we have looked at include:

- the use of true/false "probe" questions to measure situational awareness and the understandability of headquarters outputs,
- the use of Organisational Network Analysis with calculation of average distances between people,
- interoperability metrics such as the OIM,
- workload metrics such as the NASA Task Load Index,
- teamwork metrics, and
- timing measures such as the time taken to react to events, the time taken to perform tasks (including the time taken to produce the final plan), and the throughput of tasks per hour.

However, the limited range of suitably evidence-based metrics highlights the need for further research in developing C2 metrics, such as in the area of coordination metrics (Peffer *et al*. 2008), as well as the need for further studies validating their use.

Table 1: Summary of Metrics Discussed.

	Metric	Validity	Feasibility
Document Metrics	Understandability – Restatement	★★	★★
	Understandability – True/False Questions	★★★	★★
	Data Sources – Number	★★	★★
	Data Sources – Timeliness	★★	★★
Process Metrics	Process – Timing	★★★	★★★
	Process – Breadth	★	★★★
	Process – Workload	★★	★★★
	Process – Teamwork	★★	★★★
	Process – Interoperability	★★	★★★
	Process – Aggregated Measures	★	★★
	Organisational Network – Average Degree	★	★★★
	Organisational Network – Average Distance	★★★	★★★
Cognitive Metrics	SA – SAGAT	★★	★
	SA – True/False Questions	★★★	★★
	SA – Team SA	★★	★★

6. References

Cheah, M., Chew, L.P., Fong, G., Teh, C.A., and Toh, E. (2005), "Command Post Anywhere Experiment – Exploiting the use of TeamSight for Ops," *Proceedings of the 10th International Command and Control Research and Technology Symposium (ICCRTS)*, www.dodccrp.org/events/10th_ICCRTS/CD/papers/214.pdf

Cheah, M. and Fong, G. (2006), "Towards realising Parallelism for Echelons of Command," *Proceedings of the 2006 Command and Control Research and Technology Symposium (CCRTS)*, www.dodccrp.org/events/2006_CCRTS/html/papers/016.pdf

Clark, T. and Moon, T. (2001), "Interoperability for Joint and Coalition Operations," *Australian Defence Force Journal*, **151**, 23–36, www.defence.gov.au/publications/dfj/adfj151.pdf

Dekker, A.H. (2006), "Revisiting 'SCUDHunt' and the Human Dimension of NCW," *Proceedings of the 11th International Command and Control Research and Technology Symposium (ICCRTS)*, www.dodccrp.org/events/11th_ICCRTS/html/papers/029.pdf

Dekker, A.H. (2007), "Studying Organisational Topology with Simple Computational Models," *Journal of Artificial Societies and Social Simulation*, **10** (4), jasss.soc.surrey.ac.uk/10/4/6.html

Desouza. K.C., Roy, S., and Lin, Y. (2008), "Performance Measures for Edge Organizations: A Preliminary Report," *Proceedings of the 13th International Command and Control Research and Technology Symposium*, www.dodccrp.org/events/13th_iccrts_2008/CD/html/papers/163.pdf

Ford, T.C., Colombi, J.M., Graham, S.R., and Jacques, D.R. (2007), "A Survey on Interoperability Measurement," *Proc. 12th International Command and Control Research and Technology Symposium*, www.dodccrp.org/events/12th_ICCRTS/CD/html/papers/096.pdf

Guderian, H. (1952), *Panzer Leader*, Penguin (English translation).

Hart, S.G. (2006), "NASA-Task Load Index (NASA-TLX); 20 Years Later," available from NASA-TLX web site at humansystems.arc.nasa.gov/groups/TLX/index.html

Jarvis, D.A. (2005), "A Methodology for Analyzing Complex Military Command and Control (C2) Networks," *Proceedings of the 10th International Command and Control Research and Technology Symposium (ICCRTS)*, www.dodccrp.org/events/10th_ICCRTS/CD/papers/099.pdf

Kruger, J. and Dunning, D. (1999), "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments," *Journal of Personality and Social Psychology*, **77** (6), 1121–1134, www.apa.org/journals/features/psp7761121.pdf

Kyne, M.M., Militello, L.G., Thordsen, M.L., and Klein, G. (2002), "Teamwork Assessment Scales for C2 Functions of Battalions and Brigades," US Army Research Institute (ARI) Research Note 2002-08, handle.dtic.mil/100.2/ADA400488

NATO Research and Technology Organisation (2002), *NATO Code of Best Practice for C2 Assessment*, www.dodccrp.org/files/NATO_COBP.pdf

Natter, M., Ockerman, J., and Baumgart, L. (2008), "Review of Cognitive Metrics for C2," *Proceedings of the 13th International Command and Control Research and Technology Symposium (ICCRTS)*, www.dodccrp.org/events/13th_iccrts_2008/CD/html/papers/061.pdf

Nofi, A.A. (2000), *Defining and Measuring Shared Situational Awareness*, Center for Naval Analyses report CRM D0002895.A1/Final, www.cna.org/documents/D0002895.A1.pdf

Peffer, J.E., Tittle, J.S., Gualtieri, J.W., Elm, W.C., Voshell, M., Prue, B., and Woods, D.D. (2008), "How costly is your C2 Coordination? Assessing the coordination requirements within Command and Control," *Proc. 13th International Command and Control Research and Technology Symposium*, www.dodccrp.org/events/13th_iccrts_2008/CD/html/papers/153.pdf

Perla, P., Markowitz, M., Nofi, A., Weuve, C., Loughran, J., and Stahl, M. (2000), *Gaming and Shared Situation Awareness*, Center for Naval Analyses report CRM D0002722.A2/Final, www.cna.org/documents/D0002722.A2.pdf

Stanners, M. and French, H.T. (2005), "An Empirical Study of the Relationship Between Situation Awareness and Decision Making," DSTO Technical Report DSTO-TR-1687, http://dspace.dsto.defence.gov.au/dspace/handle/1947/4318

Team Management Services (2008), *Star Performing Team Profile*, www.tms.co.nz/webpages/products/spt.htm

Thomas, J.A., Pierce, L.G., Dixon, M.W., and Fong, G. (2007), "Interpreting Commander's Intent: Do we really know what we know and what we don't know?" *Proceedings of the 12th International Command and Control Research and Technology Symposium (ICCRTS)*, www.dodccrp.org/events/12th_ICCRTS/CD/html/papers/183.pdf

Wasserman, S. and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press.

US Department of Defense (2004), *Joint Command and Control Functional Concept*.