Measuring Crew Resource Management: Challenges and Recommendations

Alison Kay, Paul M. Liston, and Sam Cromie

Centre for Innovative Human Systems, School of Psychology, Trinity College, University of Dublin, Ireland
{alison.kay,pliston,sdcromie}@tcd.ie

Abstract. This paper presents a methodology for measuring Crew Resource Management (CRM) parameters as applied to a pilot decision-making task. Six teams of pilots took part in a desk-top decision-making exercise. Flight crew performance was observed by human factors researchers and was measured on a number of parameters pertaining to communication, situational awareness, decision-making, mission analysis, leadership, adaptability and assertiveness. This methodology facilitated the mapping of decisions in the context of the overall process. The communication analysis can be considered more objective than standard CRM expert rating. This methodology could be used to examine CRM for training, recruitment, incident and accident analysis, identifying degraded performance on the flight-deck and has further implications for multi-team co-ordination. It could also be used to provide a sound contribution to the design of automatic means of detection for CRM metrics on the flight deck.

Keywords: teamwork, communication, crew resource management.

1 Introduction

Crew Resource Management was first introduced in the 1980's and has since moved from the world of aviation into other sectors such as healthcare, rail and maritime industries. Good CRM is essential if safe practices are to be upheld regardless of industrial application. CRM research and application in industrial settings have progressed considerably over the last 30 years. Culture changes within organisations over the years and the acceptability of CRM in the workplace has led to CRM being "considered to be a way of working life and it is considered a definitive fact (and is now assumed) that humans do and will make errors and that good CRM is fundamental to recovering from those errors, for managing threats, risks and errors when they present themselves." (Harris, 2011).

2 Challenges for Effective CRM on the Flight-Deck

CRM has been in place in aviation for 30 years and thus is not a new concept. If aviation is such a safe industry, can CRM add anything new to flight deck

D. Harris (Ed.): EPCE 2014, LNAI 8532, pp. 480–490, 2014.

operations? Could its proliferation create the danger of CRM fatigue? The salient points of CRM have also been transferred into other industries such as healthcare, nuclear, other transport industries, chemical, petrochemical and the process industries with increasing success (Hayward and Lowe, 2010). The civil aviation authority carried out an evaluation of CRM in the UK a number of years ago (CAA, 2003). This evaluation report recommended that the content for single pilot CRM training be examined. This is ever more prescient given that flights may be operated by single crew who are supported from the ground in the not too distant future. How will CRM be affected with these anticipated changes to reduced crew and further increases in automation on the flight deck? Automatic monitoring of CRM could be used to anticipate changes in pilot performance and assist in diagnosing gradually changing levels of incapacitation. Further culture changes in CRM application within organisations are likely if CRM is to be monitored and trained for between ground stations and remotely supported aircraft operators. The research reported herein addresses these questions. Its purpose was to examine CRM metrics as applied to a decision-making task. Human factors researchers analysed CRM parameters within the context of a decision-making task in order to establish whether viewing CRM from multiple angles could give a comprehensive picture of the parameters mapped within the overall process and if this type of picture could then be applied to the operation on the flight deck.

3 Methodology

The validation methodology was based upon previous research which examined distributed situational awareness in a command and control environment (Stewart et al., 2008 and Kay et al., 2008). The methods used were observations, a modified Social Network Analysis (SNA), Hierarchical Task Analysis (HTA), Process Mapping, Co-ordination Demand Analysis (CDA) and Triangulation.

3.1 Observations

Two researchers took part in each data collection phase - both of whom were trained in the collection method for SNA. Researchers positioned themselves near pilots so that they could observe communication. They synchronised their timing devices and made note (using pen and paper) of every instance of communication such as verbal communication, head nod, hand gesture, pointing at the screen. The start and end points for data collection were agreed prior to each data collection session. The raw data from observations was put into electronic format for use in further analyses.

3.2 Modified Social Network Analysis (Communication Counts)

SNA provides a visual and numerical picture of the communication between people. It is used to analyse and represent the peoples' relationships and describes them in terms of how often they communicate, how important people seem to be within a network and how close they may be in the network. This is of interest to this research because it not only gives a visual representation of what the communication is like, but a numerical count of how much communication is taking place. It would not be typical to use SNA for teams of two people as there would generally be a challenge and response nature to the communication (i.e. if one person asks a question, the other person is likely to respond with an answer. Pilots are obliged to communicate in this way on the flight deck). Instead of having the typical network diagram showing multiple people in the network, there would be a figure showing a two people connected by one arrow. This will not provide enough information to make any inference about CRM performance, however, when communication count data is supplemented with information from the process maps a much deeper analysis can be carried out. Being able to comment on the communication frequency between pilots for specific tasks and decision points and being able to determine how information is passed (e.g. verbal commentary, hand signals, written word, and electronic messages) and how this contributes to individual and team contributions to the mission could be invaluable in creating an accurate representation of effective communication and teamwork on the flight-deck.

3.3 Co-ordination Demand Analysis (CDA)

A Hierarchical Task Analysis is carried out as the first step of the CDA. The purpose of the HTA in this research was to provide detailed task information required to feed both the CDA and the process maps. The HTA details the goals and step-by-step tasks involved in a process from start to finish. Each task and subtask within the HTA is classified as either related to task or teamwork. Teamwork related activities are given a rating for each of the teamwork taxonomy criteria. CDA produces a value for the tasks carried out in relation to the total task work, total teamwork and the levels of co-ordination between pilots.

Figure 1 (below) highlights the curricula recommendations for CRM training. The elements listed in both the JAA and FAA recommendations are in keeping with the metrics examined for CDA which are: Communication, Situational Awareness, Decision-making, Mission Analysis, Leadership, Adaptability and Assertiveness (Burke, 2005).

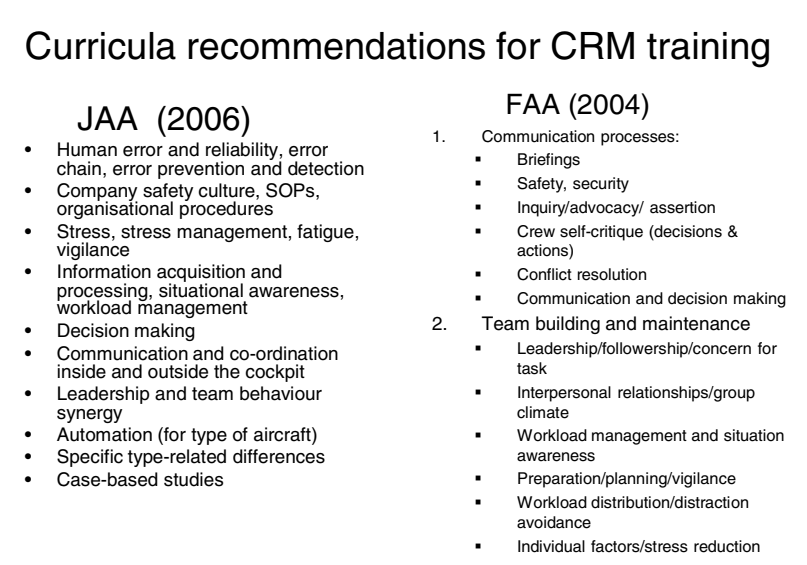

Fig. 1. Curricula Recommendations for CRM training

The commonality of metrics between the curricula recommendations and the current industry standards for measuring CRM and CDA criteria was considered sufficient for CDA to be justifiable as a means with which to examine CRM.

3.4 Process Mapping

A process map for each session was developed. In essence, the process map is a visual representation of the task analysis with a greater level of detail on the people involved in the activities, their resources, flows of information and timelines. The breakdown of tasks and goals that exist within the HTA is mapped out in a systematic manner including the logic and detail of the HTA in conjunction with the visual representation of the communication picture presented in the SNA. This allows researchers to map tasks, subtasks decision points and communication patterns to specific parts of the session. Process mapping affords greater inference of results than an HTA would in linking procedures and processes with individual tasks, personnel, communication and co-ordination. Researchers can then determine how well tasks have been achieved, information has been communicated and how all of this relates to the overall success of the session.

3.5 Triangulation

Researchers can build up a rich picture of each pilot session with SNA – communication counts, decision mapping within the process maps and CDA analysis.

Further triangulation is carried out in the processing of raw data. Transcripts were drawn up and coding applied to all the decision points for each session (i.e. three human factors researchers examined the transcripts and recordings and verified the accuracy of the coding). Coding the transcripts afforded pinpointing of the decision points in the process map and allowed decisions regarding each item being selected to be highlighted to the speech analysts so that they were able to identify which part of the recordings to examine.

3.6 Coding

Recordings from observations are transcribed and coded in order to prepare the data for further analyses using the methods used in this research. The transcriptions are coded to highlight decision points. An example from this research is shown in Figure 2 (below). Each of the 15 items were highlighted a different colour so that they could be easily identified within the text of the transcription. Beside each instance of an item within the transcription, the following reference points and times are noted: 1) the first mention of each item, 2) each decision point made 3) every review of each decision point, 4) the final decision point for each item.

Speaker A: "I think we should definitely take the water (I) (2.17). We don't need a mobile phone (I)(2.18) – there wouldn't be any reception out there anyway. Yes, the water should be first. Do you agree?"

Speaker B: "Absolutely, yes. Water (D1)(2.25) first. The flare (I)(2.26)is pretty important too. What do you think? The radio (I)(2.28)is also important if we hoped to be able to contact people. Is it just a receiver? Maybe we should choose the radio(2.33) first? The water(R1)(2.35) would keep us going for a while, but isn't it more important that people know that we are out there? I think we should go with the radio – would you agree?

Speaker A: "No, I think it just looks like a normal receiver. I don't think that we'd be able to contact anyone using it. I'd say we'd be best going for the water(R2) (2.48) first."

Speaker B: "If it's only a receiver, there's no way it should come before water. Let's go with water(DF)(2.54) first.

Fig. 2. Example of coded transcription

Coding transcriptions enabled researchers to accurately map the decision points into the process map. It also enabled them to examine how decisions are made throughout each session and the impact that external factors (such as the task interference variable) may have had on CRM performance. It would then be possible to see how quickly pilots made decisions, how many times decisions were reviewed and how this related to mission success and the communication frequency outlined in the SNA. These are all fundamental to the overall analyses.

4 Empirical Work

4.1 The Experimental Set-Up

Twelve pilots took part in desktop-based decision-making task which was adapted from The Annual Handbook for Group Facilitators (1975). All pilots were volunteers. Pilots were assigned to teams of two as would naturally occur on the flight deck. They were given a briefing which informed them of the research project, the task that they would be doing, that their speech was going to be recorded and that there would be three observers (2 human factors researchers and 1 speech researcher) in the room. They were also informed that they were taking part on a voluntary basis and could stop at any time. Researchers informed pilots that all data from the trial would be stored safely, made anonymous and that they were free to ask for their data to be withdrawn at any time until the data had been pooled. Pilots were asked to sign consent forms to take part and were given contact details for the researchers should they require further information. They were then shown into the room, (the layout for which is shown in Figure 3)

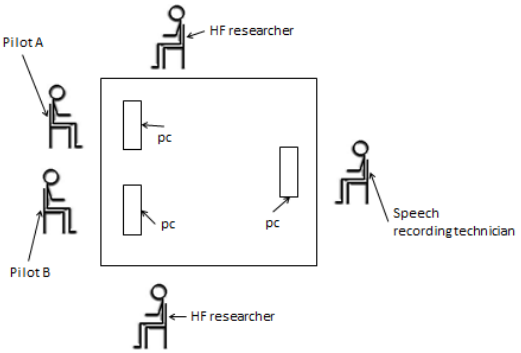

Fig. 3. Diagram of experimental set-up

Pilots had a microphone attached to their lapels and were given headsets through which they were able to hear the "beep" of an incorrect answer. Pilots were given a briefing sheet informing them they are friends who were on a ship that has sunk. Both managed to survive and are now stranded in a life boat in the middle of the sea and are at least 300 miles from the nearest landmass. 15 items were salvaged from the ship before it sank. Pilots were informed that they needed to rank these 15 items in order of importance (starting from the most important item for survival) and told to remember their choices for the session which would last for 10 minutes. All 15 items were displayed on the screen in front of them. Once items had been selected, they remained on the screen. A distractor was introduced to the task. This consisted of a beep in the pilots' headphones which could be interpreted that they had made a

mistake. This beep was generated by one of the researchers. When the 10 minutes were over the pilots were thanked for their time and researchers reiterated that they could be contacted for further information should it be required. Human factors researchers observed the session and kept counts of all types of communication. Speech analysts were present to monitor the recording and to administer the distractor.

4.2 Results

Table 1 (below) shows the pilot group performance for number of items correct and global CDA score ranking.

Table 1. Pilot group performance

Pilot Group	# items	Rank CDA
F	15	1
C	13	2
E	13	3
D	8	4
A	12	5
B	2	6

Pilot group F was the most successful at the decision making task. They also had the highest CDA score. Pilot groups C and E both had the same score for guessing the correct number of items, however Pilot group C had a higher CDA score, thus came 2nd. Pilot Group E came third. Pilot group D came fourth with a higher CDA score than Pilot groups A and B which were 5th and 6th respectively.

Group F had the highest CDA score and the greatest number of items correct. They also had significantly fewer communication counts than all of the other groups and the lowest number of incorrect answers (i.e. beeps). The CDA and decision maps showed that this group also had the lowest number of task steps of the six groups. This group completed the task in 7 minutes whilst all other groups were told that there session was over before they had completed the task.

Group C had the second highest number of items correct and the second highest number of communication. This group experienced 10 “beeps”, most from 8 minutes onwards. The time pressure and distractor had a marked effect on this group’s performance and the last 2 minutes of the session largely consisted of both pilots shouting out random items. There was little continuous dialogue from this point until the end of the session.

Group E also had the second highest number of items correct. Interestingly, this group had the second highest number of “beeps” (16 “beeps”) and the highest number

of communication counts between pilots. Inference could be made that pilots were mitigating for the effect of so many apparent "wrong" answers (i.e. indicated by each "beep"). This would have to be tested as a variable with a larger sample to draw definitive conclusions on this.

Group D had the 4th highest CDA score which may not be reflected in the number of items they got correct (8/15). Their decision-making was very good at the beginning of the session until the distractor "beep" started. This group had the greatest number of "beeps" (22) which seems to have created a great deal of stress. They also had a high communication count which may have mitigated the effect of the distractor. As mentioned for Group E, this would have to be repeated with a larger sample.

Group A came fifth in CDA score. They scored relatively highly (12/15 items) on the number of items correct, however, this group was the most affected by the distractor "beep". There does not seem to be the same mitigating factor of higher levels of communication as in groups C and E. From 7 minutes onwards, pilots seemed to have frozen and there was little or no communication from this point onwards.

Group B had the lowest score for both CDA and the number of items correct (2/15 items). This group displayed excellent prioritisation of items at the beginning of the task. Unfortunately, this was not followed through with decision making. Pilots demonstrated very thorough logic and reasoning of items but did not follow through with decisions on them.

5 Discussion

Without the communication data from the adapted SNA and the ability to examine decisions on the 15 items throughout the whole process of the decision-making task, the results would be lacking considerable detail. The adapted SNA could be considered less subjective in nature than CRM expert rating of communication as it provides a numerical value for the communication that has taken place. There was also more than one rater for each session (thus increasing inter-rater reliability). This data is mapped directly into the overall process and thus provides a more structured framework for further analyses. It could however be argued that the raters were not qualified CRM experts. Criticism levied at non-trained CRM evaluators generally concerns the understanding of the concepts behind the CRM parameters. The raters in this research are experienced human factors researchers with more than 20 years research experience in the aviation industry between them. If any criticism were to be applied here, it could be that they are both non-pilots, however, the task was a non – aviation based one, therefore this criticism is redundant. For future research applying this methodology to the flight-deck, the data will be compared to that of CRM trainers' interpretation of training sessions and will be examined by subject matter experts in incident and accident analyses. This will thus validate the data from an operational perspective.

As mentioned previously, the task used for this research was a non-aviation based one. This had merit for removing the pilots from their flight-deck environment, however, the behavior exhibited may not have been a true reflection of how pilots would have behaved for an aviation-based decision-making task. This is why the next step for the research is apply the methodology directly to behavior on the flight-deck. Interestingly, for each pilot team, the more senior of the pilots took the left hand seat in the room as would be generally represented by the pilot flying or more senior pilot on the flight deck.

The CDA rating scale had a tendency to push the researchers to choose the middle value. If the parameter under scrutiny was not extreme in nature (i.e. '1' or '3'), the researchers were forced to choose '2'. Researchers considered that it would be beneficial for the rating scale to be increased to 5, so that it is possible to give more detail for the performance on the CRM parameter. It would be useful to be able to say that performance was high (5), above average (4), reasonable (3), just under par (2), poor (1). No indication of this was possible using the current rating scale. It is therefore possible that ratings were pushed into the "medium" performance range when it could have been described otherwise. Researchers will adopt this change in rating scale in the next round of research.

The methodological approach from numerous angles make this innovation more powerful in providing a rich picture of CRM mapped into the decision-making task. Internal processes such situational awareness are difficult to assess well. Decision making was somewhat easier to analyse for this research as there was a concrete output attached to decision making in the lost at sea task. These can be clearly identified and mapped into the overall process. Situational Awareness has been shown to be difficult to accurately measure (Kirlik and Strauss, 2006, Pew, 2000, Salmon et al 2006, Stanton et al 2009). This has also been true of the measure of situational awareness in this study which had good face and construct validity but suffered from poor concurrent and predictive validity. Further research to be carried out using this methodology should also accommodate for additional measures such as a freeze technique (e.g. the Situational Awareness Global Assessment Technique) for triangulation specific to situational awareness (as evident in Stanton et al 2009,). Salmon et al 2006 recommend the use of a toolkit in measuring situational awareness. Such toolkits should include 1) performance measures, 2) a freeze probe technique, 3) a post-trial subjective rating scale and 4) an observer rating. This type of toolkit is not unlike the methodology used here.

5.1 Limitations and Further Research

The small sample size in this research has limited the amount of inference that could be made but it is hoped that further studies of simulator sessions on CRM training and live flights will provide a larger data set. The rating scale for CDA is somewhat narrow. There may be a tendency for researchers to choose the middle value. A review of the rating scale will be considered for the next round of analysis. This research did not include parameters for threat and error management. This will also be included for future research.

Contribution to Future CRM Practice. The methodological approach proposed herein and for further research differs from the original HFIDTC study (Stewart et al, Kay et al 2008) in three main areas:1) It encompasses a different systems-approach to task analysis and task modelling. 2) This research will be linked to analyses of CRM criteria from incident data and cockpit voice recordings of selected scenarios available in the public domain. Unfortunately, within the one year lifetime of this research project, but is currently being addressed. 3) In addition to criteria listed in Table 1, Loukopoulos, Dismukes & Barshi (2009) recommend that CRM training be extended for concurrent task management. This will be added as an additional criterion within the teamwork taxonomy. Concurrent task management is a concept which is extremely hard to identify using traditional task modelling and analysis tools. The use of process maps facilitates the identification of concepts such as concurrent task management. This methodology will be applied to simulated flights carried out as part of pilots' CRM training. This data will be compared to the CRM trainers' rating for the sessions. The methodology will also be applied to live flights with multiple teams on board which will facilitate analyses for within and between teams. This work will be evaluated with subject matter experts in CRM training and accident and incident analysis. Thus, the methodology will be used for the analysis of both normal and non-normal operations.

6 Conclusions

Due to the labour intensive nature of this methodology, it is unlikely that it would be employed as standard within organisations, however, it would be of great use in establishing how and where the parameters for CRM fit into the overall process of a flight or mission. This approach would facilitate the design of training for new CRM practices, especially between flight crew in the air and those on the ground for remote ground support. The methods used herein would also be useful in mapping out the gradual decline in performance from task, communication and decision-making perspectives. This would be critical to be able to further identify and define aspects of gradual incapacitation of flight-crew. This research has demonstrated that CRM parameters and decision making can be mapped into the overall process. Incapacitation is generally a gradual process. It is also very difficult to examine incapacitation in applied research settings, thus it would be very useful to be able to be able to map times throughout the flight phase that assistance would be needed from ground support and to be able to back this up using evidence where it was shown that specific CRM metrics suffered at particular points in time. There are many psychophysiological measures taken in measuring pilot performance linked to incapacitation (e.g. galvanic skin response, body temperature, heart rate, eye-tracking). Being able to back these measurements with specific behavioural measures mapped into the overall flight operational process is fundamental to being able to understand a rich picture of what gradual incapacitation looks like if there it is hoped that we will be able to automatically detect it happening on flight decks in the future. The methodology used herein could support such analyses.

In conclusion, this methodology examines CRM along similar parameters to those of LOSA and LOFT (with the exception of threat and error management), but also includes more objective measures for communication analyses. The CRM parameters and decision points can be mapped into the overall operations process, so that patterns and clusters of communication and activity can be examined in greater detail. This research has been innovative in its approach to the measurement of CRM metrics. There has been considerable effort to examine CRM metrics from several angles (SNA, CDA, Process Mapping) and the contribution of more objective measurement (communication) has been invaluable.

References

1. Burke, S.C.: Team Task Analysis. In: Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C., Jenkins, D.P. (eds.) Human Factors Methods: A Practical Guide for Engineering and Design, Ashgate, Hampshire, UK, pp. 56.1 – 56.8 (2005)
2. CAA, Crew Resource Management (CRM) Training: Guidance for flight crew, CRM instructors (CRMIS) and CRM Instructor Examiners (CRMIES) (CAP 737). Civil Aviation Authority, London (2006)
3. CAA, Methods used to Evaluate the Effectiveness of Flightcrew CRM Training in the UK Aviation Industry, CAA Paper 2002/05. Civil Aviation Authority, London (2003)
4. Flin, R., O'Connor, P., Crichton, M.: Safety at the Sharp End: A guide to non-technical skills, Ashgate, Surrey, UK (2008)
5. Harris, D.: Human Performance on the Flight Deck, Ashgate, Surrey, UK (2011)
6. Hayward, B.J., Lowe, A.R.: The Migration of Crew Resource Management Training. In: Kanki, B., Helmreich, R., Anca, J, eds. (2010) Crew Resource Management, ch. 12, 2nd edn., Wiley, San Diego (2010)
7. Kanki, B., Helmreich, R., Anca, J.: Crew Resource Management, 2nd edn. Wiley, San Diego (2010); Kay, A., Lowe, M., Salmon, P.S., Stewart, R., Tatlock, K., Wells, L.: Case Study in RAF Boeing E3D Sentry. In: Stanton, N.A., Baber, C., Harris, D, eds. Modelling command and Control, Ashgate, Surrey (2008)
8. Kirlik, A., Strauss, R.: Situation awareness as judgment I: Statistical modeling and quantitative measurement International Journal of Industrial Ergonomics 36 (2006)
9. Pew, R.W.: The state of situation awareness measurement: Heading toward the next century. In: Endsley, M.R., Garland, D.J. (eds.) Situation Awareness Analysis and Measurement, pp. 33–50. Erlbaum, Mahwah (2000)
10. Salmon, P.M., Stanton, N.A., Walker, G.H., Jenkins, D., Ladva, D., Rafferty, L., Young, M.: Measuring situation awareness in complex systems: comparison of measures study. International Journal of Industrial Ergonomics 39(3), 490–500 (2009)
11. Salmon, P.M., Stanton, N.A., Walker, G., Green, D.: Situation awareness measurement: A review of applicability for C4i environments. Applied Ergonomics 37, 225–238 (2006)
12. Stanton, N.A., Baber, C., Harris, D.: Modelling command and Control: Event Analysis of Systemic Teamwork, Ashgate, Surrey, UK (2008)
13. Stewart, R.J., Stanton, N.A., Harris, D., Baber, C., Salmon, P., Mock, M., Tatlock, K., Wells, L., Kay, A.: Distributed situational awareness in an airborne warning and control aircraft: application of a novel ergonomics methodology. Cognition Technology and Work (10), 221–229 (2008)
14. Pfeiffer, J.W., Jones, J.E.: The 1975 Annual Handbook for Group Facilitators. University Associates, Incorporated, La Jolla (1975)