Assessing team focused behaviors in emergency response teams using the shared priorities measure

Peter Berggren
Swedish Defence Research Agency
peter.berggren@foi.se

Björn JE Johansson
Linköping University, Sweden
bjorn.j.johansson@liu.se

Nicoletta Baroutsi
Linköping University, Sweden
nican529@student.liu.se

Isabelle Turcotte
Laval University, Canada
isabelle.turcotte.2@ulaval.ca

Sébastien Tremblay
Laval University, Canada
Sebastien.Tremblay@psy.ulaval.ca

ABSTRACT

The purpose of this work in progress paper is to report on the method development of the Shared Priorities measure to include content analysis, as a way of gaining a deeper understanding of team work in crisis/emergency response. An experiment is reported where the performance of six trained teams is compared with the performance of six non-trained teams. The experiment was performed using an emergency response microworld simulation with a forest fire scenario. Dependent measures were simulation performance, the Crew Awareness Rating Scale (CARS), and content analysis. Trained teams performed better and scored higher on measures of team behaviors.

Keywords

Team behavior, performance assessment, team situation awareness, shared priorities, microworld studies, CARS, method development

INTRODUCTION

Teamwork is defined as activities where a number of persons perform specific tasks and/or roles to achieve goals (Orasanu & Salas, 1993). The organization and coordination of teams affects function and performance (Jones & Roelfsma, 2000; Rasker, Post & Schraegen, 2000). Crisis and emergency response efforts are mostly team based, especially in the initial phases of a critical event. Typically, teams of responders are sent to an area where they have to assess the situation and initiate various efforts based on own initiative. As the response effort evolves, the organization of the response effort mostly stabilize and most tasks become routine, or at least planned in advanced. Work is still performed in teams at these stages, but work is orchestrated to a larger extent. When facing the initial, chaotic phase, of a response effort, it is crucial that teams are able to establish a common understanding of what is happening, what they consider important, what tasks that should be prioritized and in what order. On an individual level, this is often referred to as "situation awareness" (SA, see for example Endsley, 1987; 1995). Several tools have been developed to measure individual SA, such as SAGAT, SART and SA SWORD (Gawron, 2008; Endsley & Garland, 2000). There are also tools for measuring situation awareness in teams (Team SA) such as the Crew Awareness Rating Scale (CARS, McGuiness & Foy, 2000). CARS include two subscales: a content subscale and a workload subscale. The content subscale is based on Endsley's (1995) model of SA. The content subscale asks the participant about the perception of elements in the current situation (level 1), comprehension of the current situation (level 2) and projection of future states of task SA requirements (level 3). However, most methods approaching the measure of team SA suffer from an important drawback, namely that they depart from either self-rating on different scales, or expert ratings of team members. The problem with this approach is that there is a risk that the ratings never actually reflect the actual team SA, but rather the feeling of cohesiveness in a team. Expert ratings can be a way of working around this problem, but instead demand a very high degree of understanding of the work being conducted. In crisis response, this is very difficult as every situation tends to be unique.

Berggren and colleagues have worked for a long time (Berggren, in press) trying to create a simple, but yet reliable way of measuring team SA/cohesiveness that avoid the pitfalls mentioned above. This measure is called Shared Priorities. The purpose of this paper is to report on the development of the Shared Priorities measure to include content analysis, as a way of gaining a deeper understanding of teamwork in crisis response.

BACKGROUND

The Shared Priorities measure is based on the idea that team members rank order items that are directly related to their current work situation. Typically, five items are used. A team member can usually perform this task within 2 minutes. The researcher then calculates Kendall's measure of concordance (Kendall, 1975), i.e. to what degree the team members have ranked the items in a similar way. The way to generate items has varied between different studies. Berggren et al. (2008) used predefined items in a study on battle tank officers perception of what was important for the tank crew to perform. In that study, all team members showed the same appreciation of the situation regardless of performance. That is, the order of the predefined items was natural to the team members and could not be used to differentiate between variations in performance. In 2010, Berggren et al. tested a different approach where the team members generated the items to be ordered. This simplified and improved the shared priorities measure since the items were generated faster, were more closely related to the situation, had higher face-validity, and also needed less preparation time. This study showed a significant effect of the shared priorities measure. Prytz et al (2010) carried out a follow-up study with teams using students as participants and team member generated items. This design showed no main effect of shared priorities. That the shared priorities results from this study (Prytz et al., 2010) differ from the former study (Berggren et al., 2010) is probably explained by the fact that student teams were less cohesive than the professional teams since they had no former experience of each other and the task.

To overcome this, another study was performed where three-member teams of students trained for a long time in a simulated emergency response task in order to minimize the risk for low cohesiveness (Baroutsi et al., 2013). To simulate an emergency response situation, a simulated forest fire fighting task was used where each team member took responsibility over a number of fire engines (Granlund, 1997; 2002; Johansson et al., 2010). A control group, which received considerably less training, was also used. Further, a development of the Shared Priorities was made by introducing a *content analysis* of the generated items. When using only Kendall´s measure of concordance, Shared Priorities only told whether or not team members ranked items in a similar way. It did not tell anything about the quality of the items, and thus the quality of teamwork. The assumption that teamwork quality would be reflected in the formulation of the items is straightforward: high performing teams would likely generate items that are more general, i.e. apply to the whole team's actions rather than an individual's actions. We thus assume that experienced teams would express a higher degree of cohesiveness, suggesting that items such as "save the houses north of X town" is a more high-level goal than "supply fire engine 3 with more water" or simply "help me". Such team-focused statements reflect team behavior (Essens et al., 2005).

In the study reported in this paper, we looked at the relationship between training (high/low), content analysis of items, performance in the simulation and the Crew Assessment Rating Scale (CARS). The following hypothesizes are suggested: 1) There is a difference between trained and non-trained teams on outcome of content analysis of items (trained teams are more likely to score higher on the content analysis), 2) teams that receive high scores on the content of analysis also perform better in the simulation, and 3) scores on the content analysis and CARS are positively correlated with each other.

METHOD

A split plot design was used, team type (trained teams vs non-trained teams) X scenario type (full view scenario vs limited view scenario). Scenario order was balanced over runs. It should be noticed that by "trained", we refer to training in the microworld environment. The participants who had participated in ten successive runs were expected to perform better in the microworld than the non-trained teams.

Participants

Twelve teams with three members in each team participated. There were 28 men and 8 women. The mean age of the participants was 28.9 years. All participants were university students with no prior experience of either emergency management or the simulated environment. There was no significant difference between trained and non-trained teams regarding age or gaming experience. The trained teams were trained for 10 scenarios that included simulator runs and after action discussions. For details, see Baroutsi et al. (2013). The non-trained teams received a brief introduction including simulator runs and organizational structure (22 minutes).

Equipment

As mentioned above, a microworld simulating a forest fire-fighting scenario was used in the experiment. Microworlds have been used for several decades and typically present the participants in a study with a situation that is complex, dynamic and opaque (Brehmer & Dörner, 1993). The software used in this study is called C3Fire (Granlund, 2002) and was specifically developed for the study of team decision-making. The participants in a C3Fire experiment are expected to take the roles of fire-fighters, controlling various fire engines, water trucks and fuel trucks that must cooperate to extinguish the forest fire. Additional priorities, such as saving different kinds of housing (schools, hospitals etc) are also part of the task. The actual simulation consists of a number of cells which can take one of the following states at a given point in time; burning, burned out, fire extinguished and non-burning. The participants direct the movement and activities of their units, affecting the state of the simulated world as they move around. All this happens in real time, forcing the participants to make decisions and act upon them as the simulated area otherwise would be devoured by fire quickly.

Measures

Dependent variables were simulation performance, CARS, and content analysis. Performance in the simulation was calculated from the number of burned out cells. Different cells had different values depending on the type of object in it. Each team's score was divided with a baseline score and then subtracted from 1, resulting in a value between 0-1 where 0 indicated poor performance and 1 indicated perfect performance.

The CARS-measure included eight questions. For every question the participant is asked to rate themselves on a 4-point Likert scale, adopted from Prytz et al. (2010). The CARS questionnaire was analyzed by calculating the mean for all CARS-questions for each team and scenario.

All team members were asked to individually generate a list of five items that were important for the team to perform in the current situation. The participants received examples of good items and were reminded that the items were meant to be on a team level. In the content analysis, all items were evaluated, i.e., for every occasion every member generated five items. The items were either on A) a team level, with a specific plan, B) team level with a general plan, C) on an individual level, or D) not relevant. The classification of the items was done by two raters. The inter-rater reliability (Cohen's kappa) was 0.62 ($p<0.001$). A kappa coefficient value over 0.6 is considered substantial agreement (Landis & Koch, 1977).

If an item fulfilled the criteria the team received one point. The assessment criteria for each item was whether it was situation specific (for example, acceptable - extinguish second fire from east side, not acceptable – put out fire) and had team focus (for example, acceptable – use fast moving trucks to scout third fire, not acceptable – I need water). Maximum points for each team and occasion hence was 15.

Procedure

After having received the desired amount of training (see above), an experiment session was performed. The participants could choose among themselves what role to operate: Fire Chief, Water Chief, or Fuel Chief. Each role controlled multiple vehicles. The participants were divided by screens for viewing obscurity reasons. The teams took part in two simulator runs in C3Fire, each taking 25 minutes. After each run was finished the participants responded to the Shared Priorities instrument and CARS. Log-files were automatically saved by the C3Fire simulation server.

RESULTS

Simulation performance was analyzed using a repeated measure ANOVA. Scenario type is the repeated measure and team type is the independent measure. A main effect was found for scenario type
$F(1, 10) = 29.63$, $p < 0.001$. The full view gave an average score of 0.60 ($sd = 0.05$) and the limited view 0.39 ($sd = 0.04$). There was also a main effect found for team type $F(1, 10) = 15.38$, $p = 0.003$. Trained teams

performed better (M = 0.65, sd = 0.06) than non-trained teams (M = 0.34, sd = 0.06). No interaction effect was found.

A repeated measure ANOVA used *CARS as dependent measure*, team type as independent measure and scenario type as repeated measure. A main effect was found for scenario type $F(1, 10) = 38.20$, $p < 0.001$. The groups achieved a higher score in the full view scenario (M = 3.16, sd = 0.48) than in the limited view scenario (M = 2.46, sd = 0.47). No main effect for team type was found, $F(1, 10) = 1.26$, n.s. There was an interaction effect of scenario type and team type $F(1, 10) = 9.02$, $p = 0.013$. In the full view scenario the trained teams rated themselves higher than the non-trained teams, but there was no difference between team types for the limited view scenario.

The content analysis was analyzed using a repeated measure ANOVA with team type as independent measure and scenario type as repeated measure. A main effect was found for team type $F(1, 10) = 10.27$, $p = 0.009$. Trained teams achieved higher scores (M = 12.17, sd = 3.01) than non-trained teams (M = 6.58, sd = 3.70). For scenario type there was no significant effect ($F(1, 10) = 4.08$, $p = 0.071$). There was no interaction effect.

All measures correlated positively, see details in Table 1.

Table 1. Table of correlations. P-value in parenthesis.

	Simulation Performance	CARS
CARS	0.54 (.007)	
Content Analysis	0.64 (.001)	0.43 (.036)

DISCUSSION

The following hypothesizes were investigated: 1) There is a difference between trained and non-trained teams on outcome of content analysis of items (trained teams are more likely to score higher on the content analysis), 2) teams that receive high scores on the content analysis also perform better in the simulation, and 3) scores on the content analysis and CARS are positively correlated with each other. As seen in the result section, no hypothesis could be rejected.

From the content analysis measure it is clear that the trained teams had a better ability than the non-trained teams to generate ranking items for the shared priorities measure that had a stronger team focus and related better to the task. This can be seen as reflection of how a team has matured. It is hard for a single team member who is fully occupied with dealing with individual tasks to raise his/her view to the "team level". Focus is on delivering your part, which makes it hard to provide back-up, team support, create and share plans, cf. team behaviors. This in turn affects overall team performance. In contrast, CARS could not distinguish between trained and non-trained teams. CARS, being an established measure of team situation awareness, could be expected to distinguish between trained and non-trained teams, as trained teams could be expected to have a better team situation awareness, but we found no evidence for this. On the other hand, CARS could detect differences between the different scenario views. It is reasonable to assume that a full view is a better basis for good situation awareness than a limited view, which probably is what is reflected in the CARS measure.

The correlation matrix indicates that content analysis score is positively correlated with simulation performance. This implies that teams that are able to maintain a higher level of team-focused behaviors perform better. CARS also showed a correlation with simulation performance, but a weaker correlation than the content analysis measure.

From a practical point of view, a strong benefit of using the shared priorities measure is the easiness of administrating it. As the team members created the ranking items themselves, and these items in themselves can be used for investigating the quality of the team, data collection becomes very easy to handle for the researcher/evaluators. Also, the uniqueness of emergency response situations is yet an argument for using this type of method as it is difficult to create pre-defined items for ranking that makes sense to the participants when the situation they are about to face per definition is unknown. It is possible to create sensible pre-defined items in a simulation (however labor consuming), but in real situations, or even exercises, this becomes increasingly difficult. Content analysis of participant-created items is therefore an argument. Performing a content analysis is helpful, as it shows not only if a team prioritizes goals in a similar fashion, but also if they have achieved a good basis for team-focused behavior. Good teams are the foundation for crisis and emergency response. To achieve good teams there is a need for sound assessment techniques that can be correlated with actual outcome of response operations.

REFERENCES

1. Baroutsi, N., Berggren, P., Nählinder, S., & Johansson, B. (2013) *Training teams to collaborative as cohesive units* (Scientific report No. FOI-R--3830--SE): FOI.
2. Berggren, P. (In press) The advance of a valid and reliable tool for assessing shared understanding. In P. Berggren, S. Nählinder & E. Svensson (Eds.), *Assessing Command and Control Effectiveness – Dealing with a changing world*. Farnham: Ashgate.
3. Berggren, P., & Johansson, B. (2010) Developing an instrument for measuring shared understanding. In *Proceedings of the 7th International Conference on Information Systems for Crisis Response and Management: Defining Crisis Management 3.0*. Seattle, WA. May 2-5.
4. Berggren, P., Svensson, J., & Hörberg, U. (2008) *Mätning av gemensam lägesbild vid ledning på stridsteknisk och taktisk nivå - Studie genomförd på TCCS* (Användarrapport No. FOI-R--2647--SE). Linköping: Informationssystem.
5. Brehmer, B. & Dörner, D. (1993). Experiments With Computer-Simulated Microworlds: Escaping Both the Narrow Straits of the Laboratory and the Deep Blue Sea of the Field Study. *Computers in Human Behaviour, 9*: 171-184.
6. Endlsey, M. R. (1995) Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors,* 37(1), 32–64.
7. Endsley, M. R., & Garland, D. (Eds.). (2000) Situation awareness analysis and measurement. Mahwah: Lawrence Earlbaum Associates, Inc. Publishers.
8. 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* (Technical report No. AC/323(HFM-087)TP/59): NATO.
9. Gawron, V., J. (2008) Human performance, workload, and situational awareness measures handbook (2nd ed.). Boca Raton, FL: CRC Press.
10. Granlund R, Johansson B, & Persson M. (2001) C3Fire a Micro-world for Collaboration Training in the ROLF environment. *In proceedings to SIMS 2001 the 42nd Conference on Simulation and Modelling, Simulation in Theory and Practice*. Organized by Scandinavian Simulation Society, Porsgrunn, Norway, 8-9 October.
11. Granlund R. (2002) Monitoring Distributed Teamwork Training. Linköping Studies in Science and Technology, Linköping University Press, Linköping.
12. Granlund, R. (1997) C3Fire – A Microworld Supporting Emergency Management Training. (Thesis no. 598), Linköping Studies in Science and Technology, Linköping University Press, Linköping.
13. Johansson, B., Trnka, J., Granlund, R. & Götmar, A. (2010). The Effect of a Geographical Information System on Performance and Communication of a Command and Control Organisation. *Int. J. of Human-Computer Interaction*, 26(2&3): 228-246.
14. Jones, P. & Roelfsma, P.H.M.P. (2000) The potential for social contextual and group biases in team decision-making: Biases, conditions and psychological mechanisms. *Ergonomics*, 43(8): 1129-1152.
15. Kendall, M., G. (1975) Rank correlation methods (Forth ed.). London: Charles Griffin & Company Ltd.
16. Landis, J.R., Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics.* 33, 159–174.
17. McGuinnes, B. & Foy, L. (2000) A subjective measure of SA: the Crew Awareness Rating Scale (CARS). *In The Human Performance, Situational Awareness an Automation Conference*. Savannah, Georgia, 16-19 Oct 2000.
18. Orasanu, J. & Salas, E. (1993) Team Decision Making in Complex Environments. In G. A. Klein, J. Orasanu, R. Calderwood & C.E. Zsambok (Eds.) *Decision making in action: Models and methods* (pp. 327-345). Norwood, NJ: Ablex Publishing.
19. Prytz, E., Berggren, P., & Johansson, B. J. E. (2010) *Performance and shared understanding in mixed C2-systems* (Scientific report No. FOI-R--3155--SE). Linköping.
20. Rasker, P.C., Post, W.M. & Schraagen, J.M.C. (2000) Effects of two types of intra-team feedback on developing a shared mental models in command and control teams. *Ergonomics*, 43(8): 1167-1189.