Presented at the Proceedings of the 20th Annual Conference
of the Canadian Nuclear Society
Montreal, QC, Canada
May 30- June 2, 1999

Doron Rabin
Human Factors Section
Atomic Energy Control Board

Evaluating Nuclear Power Plant Crew Performance During Emergency Response Drills

Abstract

The Atomic Energy Control Board (AECB) is responsible for the regulation of the health, safety and environmental consequences of nuclear activities in Canada. Recently, the Human Factors Specialists of the AECB have become involved in the assessment of emergency preparedness and emergency response at nuclear facilities. One key contribution to existing AECB methodology is the introduction of Behaviourally Anchored Rating Scales (BARS) to measure crew interaction skills during emergency response drills. This report presents results of an on-going pilot study to determine if the BARS provide a reliable and valid means of rating the key dimensions of communications, openness, task coordination and adaptability under simulated emergency circumstances. To date, the objectivity of the BARS is supported by good inter-rater reliability while the validity of the BARS is supported by the agreement between ratings of crew interaction and qualitative and quantitave observations of crew performance.
Keywords: human factors, crew interaction skills, emergency response

Introduction

It is widely accepted that the control room of a nuclear power plant (NPP) acts as the nerve centre for the entire nuclear power plant and that control room staff play a central role in mitigating emergency situations. Traditionally, preparation for duty in the control room has focused on mastery of formal written procedures to support rule-based performance [1]. However, there is a growing awareness that knowledge-based performance is particularly important during "non-routine" or "unanticipated" emergencies [1, 2]. In a complex NPP system, the cognitive demands of situation assessment and response planning exceeds individual ability. As a result a crew of interdependent individuals is required to combine expertise and knowledge of the system in the common goal of restoring the facility to normal operating parameters. Team decision making theories such as shared mental models or team mind suggest that a control room crew would rely on a shared understanding of facility operations and the current state of the facility to find a solution to an emergency [3, 4]. This shared understanding implies that crew members must be able to effectively interpret and communicate information amongst themselves in order to support good technical performance [5].

Indeed, there is a growing body of evidence that supports the role of crew interaction skills, such as communication, cooperation and coordination in team effectiveness [6]. For example, in 1986 Foushee, Lauber, Baetge and Acomb found that differences in group communication processes accounted for performance differences between technically qualified flight crews during cognitively demanding flight simulator trials [7]. To date, case studies of past events and applied research by commercial, military and aerospace aviation organizations have supplied the bulk of evidence for the relationship between crew interaction and decision making [for example see 6, 8]. In comparison to the aviation industry, the nuclear industry has lagged behind in assessing this potentially important aspect of facility safety.

USNRC Studies into Crew Interaction Skills

In the early 1990s, the US Nuclear Regulatory Commission began to develop and evaluate a series of tools designed to measure crew interaction skills in the context on nuclear power plant control room operations. Findings from previous research into team performance appraisal and a workshop with contract license examiners identified 7 team performance dimensions. Draft Behaviourally Anchored Ratings Scales (BARS) were developed for each dimension. Each BARS provided examples of "high", "average" and "low" team performance which were mapped on to a 7-point numerical rating scale. An initial pilot test where 11 crews were rated during simulator trials, revealed low inter-rater reliability for the BARS [9]. Subsequently, information from statistical analyses and rater comments was used to eliminate 3 of the team performance dimensions, while the remaining 4 were reworded and 2 new dimensions were added [8]. The 6 current dimensions are communications, task coordination, team spirit, maintaining task focus in transition, openness and adaptability. These dimensions, as defined in NUREG/CR-6208 are presented in Table 1 [2]:

Table 1: Definitions of the 6 Dimensions of Crew Interaction Skills

Dimension	Definition
Communications	Transaction of factual information in a clear and precise manner. In terms of crew behaviours this includes listening skills, nonverbal behaviour, and articulation of plant status or instructions for future activities. Communication does not include emotional aspects of information transmission.
Task Coordination	Crew members' ability to match the available resources, such as people and procedures, to the task in order achieve the optimal workload distribution.
Team Spirit	Mutual support, cohesiveness, group identity, and the extra effort that crew members exhibit to accomplish a common goal.
Maintaining Task Focus in Transition	Crew members' responses to changes from normal to non-normal plant conditions.

Table 1 continued

Openness	Crew members' tendency to ask for, give and receive suggestions. It includes questioning decisions and discussing alternatives to arrive at the best possible decision. Openness also involves the reactions of crew members to feedback, which should focus on task rather than on the person when reviewing actions.
Adaptability	Crew members' ability to adjust or modify their behaviour based on the situation, to be flexible in responding to the environment, and to recognize the need for change.

Adapted with permission from Roth, E.M., Mumaw, R. J. & Lewis, P.M. (1994). An empirical investigation of operator performance in cognitively demanding simulated emergencies. Prepared for the U.S. Nuclear Regulatory Commission. (NUREG/CR-6208).

The 7-point scale of the BARS was maintained, while new behavioural anchors were rewritten for each of the 6 dimensions. Thirteen observers used the BARS to rate the crew interaction skills of 8 crews responding to 3 different simulated plant emergencies. The average inter-rater reliability for the BARS was 0.89 while repeated use of the scales and training seemed to improve rater accuracy [8].

In 1994, the USNRC used the BARS as one part of a larger assessment of crew performance during cognitively demanding simulated emergencies [2]. The authors reported variability in the crew ratings for communications, openness, task coordination and adaptability. Furthermore, a comparison between crews exhibiting good technical performance and crews exhibiting poor technical performance revealed statistically significant differences on ratings of communication, task coordination and adaptation [2]. These findings provide initial support for the utility of the BARS in evaluating between-crew differences in crew interaction and indicate a possible link between team interaction skills and technical performance.

The Ongoing AECB Pilot Study

As a regulatory agency, the AECB is responsible for the regulation of the health, safety and environmental consequences of nuclear activities in Canada. Fulfilling this responsibility requires in part, that AECB staff assess emergency preparedness (EP) and emergency response (ER) programs at licensee facilities. This is accomplished by a number of co-coordinated activities each of which contributes to a multidisciplinary evaluation of EP and ER programs. In response to mounting evidence of a link between team interaction skills and technical performance [for example see 2], the AECB began to test a set of BARS developed by the USNRC for the purpose of evaluating overall crew interaction skills during emergency drill scenarios.

The Human Factors Specialists of the AECB have been testing the utility of the BARS developed by the USNRC for the purpose of evaluating overall crew interaction skills during emergency drill scenarios. In order to be considered a useful regulatory tool, the BARS must demonstrate 4 essential characteristics; broad-based utility, objectivity, quantitative and qualitative information and high scrutiny and use [10].

Broad-Based Utility: The BARS must be capable of broad based use across a variety of scenarios and crews. Ideally, the BARS should be applicable to any of the crews involved in emergency mitigation, such as control room staff or emergency operations centre staff.

Objectivity: The BARS must also provide an objective measure of performance to ensure agreement on the data collected. High inter-rater reliability and convergence with other evaluative methods is a necessary (but not sufficient) condition for the demonstration of validity. High objectivity would support conclusions based on the data collected and allow for cross-sectional and longitudinal between-crew and between-scenario comparisons.

Quantitative and Qualitative Information: Both quantitative and qualitative information is essential to a thorough evaluation of crew performance. A numerical scale provides an easy means of comparing individual evaluations and assessing objectivity. Qualitative behavioural descriptive statements provide a link between observed crew behaviours and the numerical ratings assigned to crew performance.

High Scrutiny and Use: Testing the BARS for broad-based utility, objectivity and quantitative and qualitative information requires using these tools in the field and subjecting them to peer scrutiny. Incorporating the BARS into the emergency response evaluation program of the AECB will require cross-divisional approval within the AECB.

Field Use

It is standard regulatory practice for an AECB evaluation team to attend emergency response drills at licensee facilities. Each member of the team is assigned to observe a specific position within the emergency response organization of the facility. Raw observations and the times at which they occur are recorded and later used to complete a checklist of recommended criteria for the evaluation of on-site nuclear power plant emergency exercises [12]. These checklists are then incorporated into a report outlining AECB findings and recommendations, a copy of which is sent to the participating licensee.

The BARS were incorporated into the AECB evaluation process late in 1998. An AECBHuman Factors Specialist trained other AECB observers to use the BARS.

Trials to Date

To date, the BARS have been used in the evaluation of one emergency response drill. After observing crew performance during an emergency response drill at a NPP, all members of an 8 member AECB evaluation team completed the BARS for communication, openness, task coordination and adaptability. Teams at 6 physically separated locations were observed; the site management centre (SMC), emergency operations centre (EOC), corporate emergency operations facility (CEOF), shift supervisors office (SSO), far boundary team (FBT) and emergency response team (ERT). The EOC and SSO were observed by two AECB observers each.

Results

The mean BARS scores across all teams were higher than the average scores reported for "Good Technical Crew Performance" in NUREG/CR-6208 [2]. The observers assigned to the SSO recorded nearly identical ratings on the BARS. The observers in the EOC recorded similar results on 2 of the BARS, but rated the task coordination BARS very differently. The results are presented in Table 2:

Table 2: Individual Observer Ratings and Mean Overall Ratings of Communication, Openness, Adaptability and Task Coordination

Observers	Location	Ratings			
		Communication	Openness	Adaptability	Task Coordination
1	SMC	5.5	4.5	5	6
2	EOC	5	6	-	3
3	EOC	4	7	5	7
4	CEOF	4	6	5	5
5	SSO	6	7	7	7
6	SSO	7	7	7	6
7	FBT	5	4	-	-
8	ERT	5	6	6	6
Mean Ratings		5.2	5.9	5.8	5.7

Discussion

Results from an initial trial at a licensee facility demonstrated both the strengths and weaknesses of the BARS. The seamless integration of the BARS into existing evaluative activities support their broad-based utility. Observers acted as raters with minimal additional training or instruction and were able to use the BARS to rate very different crews in the SMC, EOC, CEOF, SSO, FBT and ERT.

Overall, the objectivity of the BARS ratings was supported by good inter-rater reliability. The most striking instance of rater disagreement appeared in the task coordination ratings of the EOC crew. Informal discussions with observers revealed that the rating of 3 was in response to one incident where a discrepancy in procedures resulted in a poor coordination of activities between the off-site emergency coordinator and an off-site survey team. This finding highlights a potential weakness of the BARS; one isolated instance of poor performance can have a strong influence over a rating of overall team performance, especially with fewer observations.

The validity of the BARS was supported by the agreement between crew interaction ratings and qualitative and quantitave observations of crew performance. The AECB team concluded that the licensee facility was able to quickly recognize an emergency situation in the station, classify the event promptly and appropriately, conduct the appropriate notification and take appropriate actions to prevent damage to equipment and keep important systems running. The licensee facility met requirements for good technical performance which suggests that the mean crew interaction ratings were appropriate.

It is important to consider the limitations of this study in interpreting the results of this trial. The most notable limitation was the relatively small sample size of seven observers. The small data set allowed mean ratings to be influenced by individual ratings. Furthermore, it makes the interpretation of statistical analysis results tenuous due to low statistical power. A possible solution to these issues would be to include a larger sample size (ie. more raters) in future trials.

Another limitation arises in the field application of the BARS. The USNRC tests [2, 8, 9] of the BARS all took place in simulated control rooms, while the AECB trial applied the BARS in both control room and non-control room settings. It is unclear if the quantitatively determined reliability value of 0.89 reported by the USNRC [8] applies to the non-control room settings. Although the BARS used in the present study were designed to assess control room crews, team interaction skills are important in any team setting. Furthermore, the use of telephones, facsimile transmissions and shared procedures may support communication and coordination amongst physically isolated teams to the point where they can be considered as one larger team.

Based on the results of this trial the AECB agreed to support the continued testing of the BARS during future emergency response drills. Future trial applications of the BARS should serve to demonstrate the strengths and limitations of these measures of crew interaction skills and clarify their role as a regulatory tool.

References

[1] Mumaw, R. J., Swatzler, D., Roth, E.M. & Thomas, W.A. (1994). Cognitive skill training for nuclear power plant operational decision making. Prepared for U.S. Nuclear Regulatory Comission. (NUREG/CR-6126).

[2] Roth, E.M., Mumaw, R. J. & Lewis, P.M. (1994). An empirical investigation of operator performance in cognitively demanding simulated emergencies. Prepared for the U.S. Nuclear Regulatory Commission. (NUREG/CR-6208).

[3] Orasanu, J. & Salas, E. (1993). Team decision making in complex environments. In Klein, G.A., Orasanu, J., Calderwood, R. & Zsambok, C.E. (Eds.) <u>Decision Making in Action: Models and Methods</u>. (pp. 327-345). Norwood, NJ; Ablex Publishing Corporation.

[4] Lehto, M.R. (1997). Decision making. In Salvendy, G. (Ed.) Handbook of Human Factors and Ergonomics, 2nd ed. (pp. 1235-1242). New York; John Wiley & Sons, Inc.

[5] Gaddy, C.D & Wachtel, J. A. (1992). Team skills training in nuclear power plant operations. In Swezey, R.W. & Salas, E. (Eds.) Teams: Their Training and Performance. (pp.379-398). Norwwod, NJ: Ablex Publishing Corporation.

[6] Brannick, M.T., Prince, A., Prince, C. & Salas, E. (1995). The measurement of team process. Human Factors, 37 (3), 641-651.

[7] Foushee, H. C. & Helmreich, R. L. (1988). Group interaction and flight crew performance. In (Eds.) Human Factors in Aviation. (pp. 189-227). New York: Academic Press, Inc.

[8] Montgomery, J., Gaddy, C., Toquam, J. (1991). Team interaction skills evaluation criteria for nuclear power plant control room operators. Proceedings of the Human Factors Society 35th Annual Meeting. Research supported by the U.S. Nuclear Regulatory Commission (Contract DE-AC06-76 RLO 1830).

[9] Montgomery, J.C. & Hauth, J.T. (1990). Team interaction skills evaluation criteria for nuclear power plant control room operators. 18th Water Reactor Safety Information Meeting. Research supported by the U.S. Nuclear Regulatory Commission (Contract DE-AC06-76 RLO 1830).

[10] Haber, S.B. & Barriere, M.T. (1998). Development of a regulatory organizational and management review method. Prepared fro the Atomic Energy Control Board (RSP-0060).

[11] Lafortune, J.F. (1997). Recommended criteria for the evaluation of on-site nuclear power plant emergency exercises. Prepared for the Atomic Energy Control Board (INFO-0667).