

A multi-level approach of evaluating crew resource management training: a laboratory-based study examining communication skills as a function of team congruence

J. Sauer^{a*}, A. Darioly^b, M. Schmid Mast^b, P.C. Schmid^b and N. Bischof^a

^a*Department of Psychology, University of Fribourg, Fribourg, Switzerland;* ^b*Department of Work and Organizational Psychology, University of Neuchatel, Neuchatel, Switzerland*

(Received 19 February 2009; final version received 12 August 2010)

The article proposes a multi-level approach for evaluating communication skills training (CST) as an important element of crew resource management (CRM) training. Within this methodological framework, the present work examined the effectiveness of CST in matching or mismatching team compositions with regard to hierarchical status and competence. There is little experimental research that evaluated the effectiveness of CRM training at multiple levels (i.e. reaction, learning, behaviour) and in teams composed of members of different status and competence. An experiment with a two (CST: with vs. without) by two (competence/hierarchical status: congruent vs. incongruent) design was carried out. A total of 64 participants were trained for 2.5 h on a simulated process control environment, with the experimental group being given 45 min of training on receptiveness and influencing skills. Prior to the 1-h experimental session, participants were assigned to two-person teams. The results showed overall support for the use of such a multi-level approach of training evaluation. Stronger positive effects of CST were found for subjective measures than for objective performance measures.

Statement of Relevance: This work provides some guidance for the use of a multi-level evaluation of CRM training. It also emphasises the need to collect objective performance data for training evaluation in addition to subjective measures with a view to gain a more accurate picture of the benefits of such training approaches.

Keywords: communication skills; crew resource management training; performance; team composition

1. Introduction

The use of teams for operating complex technical work environments is quite common in many application areas. These teams often have a hierarchical structure (e.g. captain and crew member on a ship's bridge). However, the team member (TM) entrusted with the leadership position may not always be the one with the highest task competence. This may refer to highly situation-specific aspects of competence (e.g. co-pilot draws right conclusion from display information while the captain does not) as well as to situation-unspecific aspects of competence (e.g. if a very experienced pilot is employed as first officer by an airline; Ginnett 1993). In both cases, a mismatch between hierarchical position and task competence bears the risk of sub-optimal team performance.

An important goal of the present study was to employ a multi-level approach for the evaluation of communication skills training (CST) as an important element of crew resource management (CRM) training. Within this methodological framework, the work aimed to investigate how hierarchically structured teams whose hierarchical organisation was either

congruent or incongruent with the distribution of competence among the TMs (congruent = high status person is the more competent one; incongruent = low status person is the more competent one) affects team outcomes (e.g. task performance, team involvement) and how CST impacts on these relations or team outcomes.

1.1. Team member competence and hierarchical status

While the competence of individual TMs makes up the competence of a team, this may not follow a simple principle (i.e. team performance equals the sum of the performance of individual TMs) but depends on the type of task (Steiner 1972). For example, for disjunctive tasks, overall team performance is determined by the best TM (e.g. solving a mathematical problem); whereas for conjunctive tasks, the worst TM determines overall team performance (e.g. speed of task completion at an assembly line). Because many team tasks are of the disjunctive nature (Littlepage 1991), such a task was used to ascertain the ecological validity of the present study. Previous

research using the same task environment as in the present study indicated that the tasks used are of a disjunctive nature (Sauer *et al.* 2006).

While team performance is determined by the best TM in disjunctive tasks, this effect may be influenced by the hierarchical status of the individual (leader vs. subordinate). There is evidence in the literature that a matching competence/status hierarchy provides more benefits than a mismatching hierarchy. For example, Katz and Kahn (1952) showed that the competence of the higher status TM is positively related to the lower status TM's satisfaction. Team morale (i.e. the ability of a team to pull together consistently in pursuit of a common goal) was higher when there was a match between team leader's competence and status (Hamblin *et al.* 1961). However, the effects of a mismatch between status and competence may not be entirely negative. For example, in organisational decision-making, low status TMs were found to increase their participation in team decision-making processes when high status TMs were less competent (Salam 1998). The interplay between status and competence is also addressed by Expectation States Theory (Berger *et al.* 1977). For example, it proposes that if the leader does not have the required task competence, he/she may be considered illegitimate and subordinates of such 'illegitimate' leaders are more likely to resist and challenge the leaders' attempts to influence the course of action (e.g. Darioly and Schmid Mast 2010). Overall, these studies suggest that the mismatch between status and competence can be a liability for teams.

Negative effects that occur as a result of a mismatch between status and competence might be less severe if team coordination processes allow for an optimal use of team resources. Modern concepts of leadership have tried to overcome the classic model of 'static' vertical leadership, which postulates the exertion of power by a specific individual across situations (Bass 1990). In teams operating complex technical systems, the situation may require that leadership behaviour is also adopted by other TMs than the formally designated leader (e.g. first officer carries out an emergency landing). This idea is conceptualised in the concepts of situated leadership (Falk 1999) or shared leadership (Pearce and Sims 2002). Research has shown that teams operating in critical situations under time pressure benefit very much from shared leadership (Perry *et al.* 1999). Therefore, the detrimental effects of an incompetent leader together with a competent subordinate might be attenuated by such horizontal leadership concepts that emphasise the importance of effective within-team communication and a more flexible adoption of leadership behaviour. The focus of the present study was in testing whether

improving communication and coordination skills of a mismatched team with CST might be beneficial and entail a more optimal use of team resources.

1.2. Crew resource management training

Due to the negative effects that status differences may have on team performance under some circumstances, measures have been conceived to reduce their possible detrimental impact. One of these measures represents CRM training (e.g. Salas *et al.* 2001). The main goal of CRM training is to provide TMs with a range of non-technical skills needed to achieve good team performance and smooth team interaction.

Although there is no standardised training programme, CRM training follows a number of principles. Core elements of CRM training include human error and reliability, company safety culture, stress management, information processing and situation awareness, decision-making, leadership and communication (Droog 2004). A chief objective of CRM training is to equip TMs with good communication skills, which include receptiveness and influencing skills.

Receptiveness involves paying attention to the ideas, comments and questions of other TMs (Dunlap and Mangold 1998). It relates to specific behaviours, such as encouraging feedback, incorporating suggestions from others into decisions and active listening. Influencing skills involve using effective interpersonal skills and appealing to other TMs' logic in order to win support for an idea or viewpoint. It relates to specific behaviours, such as using tact when asserting a position and employing an appropriate level of assertion.

While CRM training has been widely employed, there have been concerns about its effectiveness (Salas *et al.* 2000). Two reviews of the literature on CRM training concluded that there is a great variety in the kind of data collected across studies (Salas *et al.* 2001, 2006). Following Kirkpatrick's (1976) distinction of different criteria for training effectiveness, the most positive evidence was found at the first level (reaction, i.e. how much did trainees like the training?). At the second level (learning, i.e. did the trainee increase knowledge or change work attitude?), positive evidence became less strong while at the third level (behaviour, i.e. was the trainee able to improve performance?) even less so. At the fourth level (results, i.e. organisational impact) hardly any evidence is available due to the dearth of studies measuring this aspect.

Based on the results of their literature reviews, Salas *et al.* (2006) argued for a need to carry out more studies measuring training effectiveness at several levels, following Kirkpatrick's distinction. Even at a

specific level (e.g. behaviour), there would be a need to measure various facets of the outcome (e.g. measuring several facets of behaviour, such as multiple task performance). Moreover, most of the studies reported were quasi-experimental studies (79.4%) and a reasonable number were *post-hoc* studies (17.6%) while hardly any experiments were carried out (2.9%).

1.3. Coping with sub-optimal working conditions

To determine the effectiveness with which team resources are employed, several aspects of operator behaviour can be measured. To gain a better understanding of how these different aspects of behaviour contribute to overall performance, Hockey's (1997) model of compensatory control mechanisms may be helpful. Developed in the context of stress research, the model argues that operators aim to maintain adequate levels of performance on high priority tasks (i.e. primary tasks), even if working conditions become increasingly difficult (e.g. onset of noise, extended task involvement). Performance maintenance is achieved by a compensatory process in which additional cognitive resources are recruited to the primary task. However, this may not be without cost and may therefore result in performance decrements in secondary tasks (e.g. lower priority tasks such as log-keeping are not carried out regularly) or in non-optimal task management strategies (e.g. reduced sampling of peripheral displays). In order to measure these forms of compensatory behaviour suggested by Hockey's model, a methodological approach is required that takes several measures of task performance (i.e. at the behavioural level in Kirkpatrick's model), with each measure covering a different type of task. This may be achieved by using a multiple-task environment, which comprises several sub-tasks to which different priorities are attached.

1.4. The present study

Based on the critical issues raised by Salas *et al.* (2006), the present study aimed to adopt a methodological approach that may provide a useful framework for future evaluations of CRM training. Within this methodological framework, team congruence was examined as an example of an unfavourable team composition, which allows any positive impact of CST to become evident more easily than under more optimal team compositions.

If the leader is not the one with the most pronounced task competence, it was expected that teams in which the status hierarchy did not match the task competence hierarchy would perform worse than teams in which both type of hierarchies matched.

Teams were therefore created with a formal hierarchy in which the individual task competences either matched the status roles or did not. Thus, congruent teams comprised a legitimate leader (task competent) and a legitimate subordinate (task incompetent) and incongruent teams comprised an illegitimate leader (task incompetent) and an illegitimate subordinate (task competent).

It was predicted that CRM training would be more helpful for incongruent teams than for congruent teams because the performance gains that can be made from improving within-team cooperation and communication in non-optimal team compositions are greater than in teams that already have a positive team composition. This refers to the relationship of CRM skills and task competencies. For example, if the task competencies of a TM are high, he or she may not need to make use of CRM skills to tap into their team mate's competencies.

In the present study, the focus of CRM training was on two critical communication skills (receptiveness and influencing skills) since covering the full content of CRM training was beyond the scope of the study. Receptiveness and influencing skills were chosen from the many elements of CRM training because they were considered to be critical in improving within-team communication. This would allow TMs to exchange information about how to coordinate task activities and how to cope best with fault states within the simulated process control environment. This adaptive coordination is particularly needed in situations of high demand (Grote *et al.* 2010). This applies in particular to situations where team composition is non-optimal due to incongruence. In the incongruent teams, there is a particular need for the competent assistant to make his/her specific technical knowledge known to the leader (i.e. influencing skills) rather than silently observing how the low-competence leader attempts to manage the system in a non-optimal way. Conversely, there is a special need for the low-competence leader to be sufficiently accessible to the suggestions offered by the assistant (i.e. receptiveness). For the congruent teams, any positive effects of CRM training were expected to be much smaller than for incongruent ones, with the possibility of even detrimental effects of CRM training being observed. This is because, in congruent teams, there is a lesser need to exchange critical information so that extensive (but unproductive) discussions in teams may divert cognitive resources from the main task (e.g. if a less competent assistant persistently tries to influence a competent leader to take a certain course of action).

An important theoretical basis for this methodological approach represented the model of Kirkpatrick (1976), which distinguishes between

different criteria for the evaluation of training approaches. Following these criteria, the present study measured various aspects of reactions, learning and behaviour. Therefore, a computer-based simulation was employed to model a complex process control environment. This simulation environment has good data-gathering facilities, which allows the measurement of different aspects of objective team behaviour, including primary and secondary task performance, information sampling and system control behaviour. In addition, TM evaluations of the process of teamwork were measured (e.g. team climate). Finally, an experimental approach was employed to compensate for the shortage of experimental research in the field of CRM training. Only an experimental approach with randomly assigning participants to CRM training or a control condition and then measuring an outcome can provide evidence as to whether CRM training is causally responsible for differences in the outcome.

In sum, the following predictions were made: (1) congruent teams would show better performance than incongruent teams, with secondary tasks being more sensitive for this effect than primary tasks; (2) CST would have a positive effect on team performance, with secondary tasks being more sensitive for this effect than primary tasks; (3) CST would improve performance more strongly for incongruent teams than for congruent teams, with secondary tasks being more sensitive for this effect than primary tasks.

2. Method

2.1. Participants

Participants were 64 males (age: mean 20.7; SD 2.15; range 18–29), recruited at different technical universities (in the French-speaking part of Switzerland) to ensure that they had a comparable understanding of complex technical systems. Half of the participants were majoring in computer science while the others pursued degrees in different natural sciences (e.g. physics, chemistry, biology, medicine). Participants were paid 50 Swiss francs (approx. €35) for their participation.

2.2. Experimental design

The study was a two (CST: with vs. without) by two (competence/hierarchical status: congruent vs. incongruent) between-subjects design. The level of analysis was the team, consisting of two TMs who did not know each other. CST involved random assignment of the teams to either the training or control condition. High and low status roles within teams were randomly assigned. In half of the teams (random assignment),

the leaders were made competent (congruent teams), whereas in the other half of the teams the assistants were made competent (incongruent teams) by providing a special training session with specific technical knowledge that enables TMs to better manage certain system faults encountered in the testing session.

2.3. Task environment: the Cabin Air Management System

The task environment, called Cabin Air Management System (CAMS), models a highly automated process control environment using the operational context of a spacecraft's life support system. As it has been described in detail elsewhere (e.g. Sauer *et al.* 2000), only a brief summary is given here.

The CAMS environment consists of five automatic controllers that maintain their corresponding system parameters (O₂, CO₂, cabin pressure, temperature and humidity) within a predefined target state. The parameters refer to subsystems that are closely coupled and hence have an effect on each other during system operation. During normal system operation, the operator monitors the performance of the automatic controllers and intervenes only in the event of a system fault (e.g. cooler failure, O₂ leak). Figure 1 shows the main interface of CAMS, which provides feedback about the operation of various system components. Flow meters indicate the flow of O₂ and N₂ at several locations in the system. The icon of the mixer valve rotates when either gas is flowing. The operation of the various subsystems (e.g. cooler, dehumidifier) is indicated by various symbols. Finally, the warning system issues an alarm if any of the parameters moves out of its safe range.

The screen manager enables the operator to call up various displays and control panels. The history display provides the operator with graphical information about changes in the levels of parameter over the last 4 min. The maintenance facility allows the operator to repair any system fault, with each repair being completed within approximately 60 s. Since no explicit feedback is provided to whether the diagnosis was correct, the operator needs to check whether the fault state has actually been rectified.

There are four typical process control tasks that are to be carried out by the operator. These are divided into primary and secondary tasks according to their priorities for overall system integrity.

- (i) System stabilisation: The first primary task was to maintain a stable system state at all times. This was achieved by monitoring the safe functioning of the automatic controllers

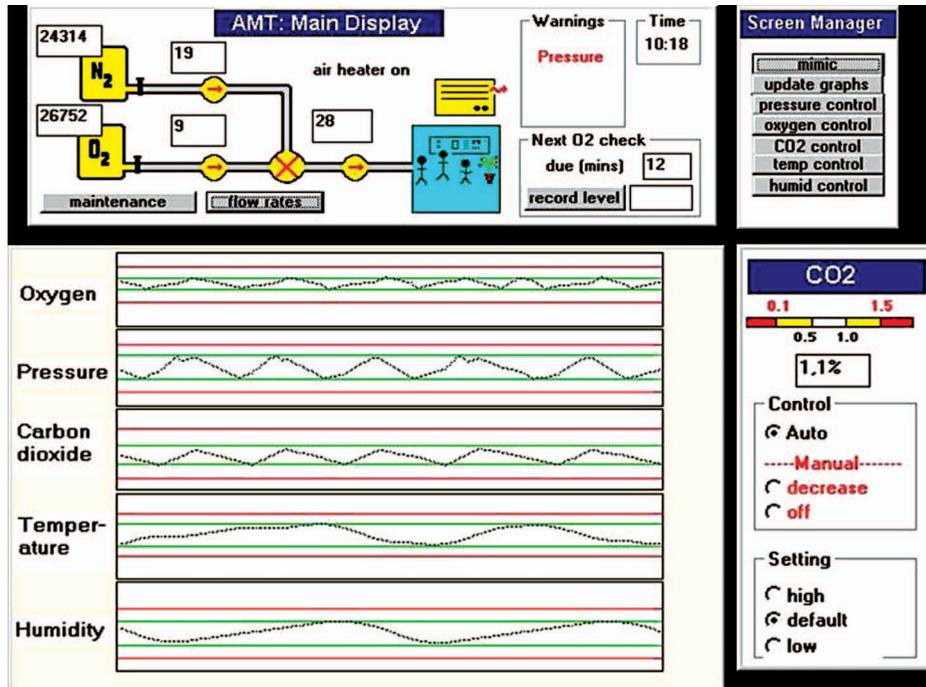


Figure 1. Main interface of Cabin Air Management System environment.

- and by adopting manual system control if required.
- (ii) Fault diagnosis: The second primary task was to identify and repair any emerging system fault.
 - (iii) Alarm acknowledgement: The first secondary task was to acknowledge any system alarm, which involved clicking on a warning signal as soon as it was displayed. This task provided a measure of reaction time.
 - (iv) Tank level recording: The other secondary task was to keep a record of O₂ tank levels by regularly checking current levels (i.e. at 3-min intervals). This corresponds to a time-based prospective memory task (i.e. to remember to complete an action at a specified time in the future).

CAMS is an appropriate simulation environment for teamwork. Although CAMS can also be operated by a single operator (because it is not a distributed system), it provides the possibility for task division between TMs, which can only be successfully achieved by means of regular within-team communication. In this way, CAMS is similar to other technical systems, in which CST is used. For example, the task of flying in civil aviation is typically carried out by a team of two pilots. In an emergency, the task could also be carried out by a single pilot (e.g. if the co-pilot is incapacitated), although this would clearly be at the cost of

excessive workload and, possibly, impaired performance. The suitability of CAMS as a team task has also been demonstrated in a previous study (Sauer *et al.* 2006). Furthermore, previous work has also demonstrated that in a variety of different sub-optimal working conditions (e.g. noise, night work), CAMS proved to be a sufficiently sensitive simulation environment (e.g. Sauer *et al.* 2003, Hockey *et al.* 2007).

2.4. Procedure

After arriving in the laboratory, participants were told that the main goal of the study was to investigate how people work together in teams managing complex technical systems. Each participant signed an informed consent form. The different elements of the procedure are summarised in Figure 2.

2.4.1. Technical training on the Cabin Air Management System and status assignment

Technical training took place in groups of six to 12 and was given by three instructors, with each of them being specialised in particular aspects of the training programme. The technical training (totalling 120 min) consisted of two phases. In the first phase (lasting about 60 min), participants received a general introduction to the operation of CAMS, using a mixture of PowerPoint presentations and practical exercises. This allowed participants to familiarise

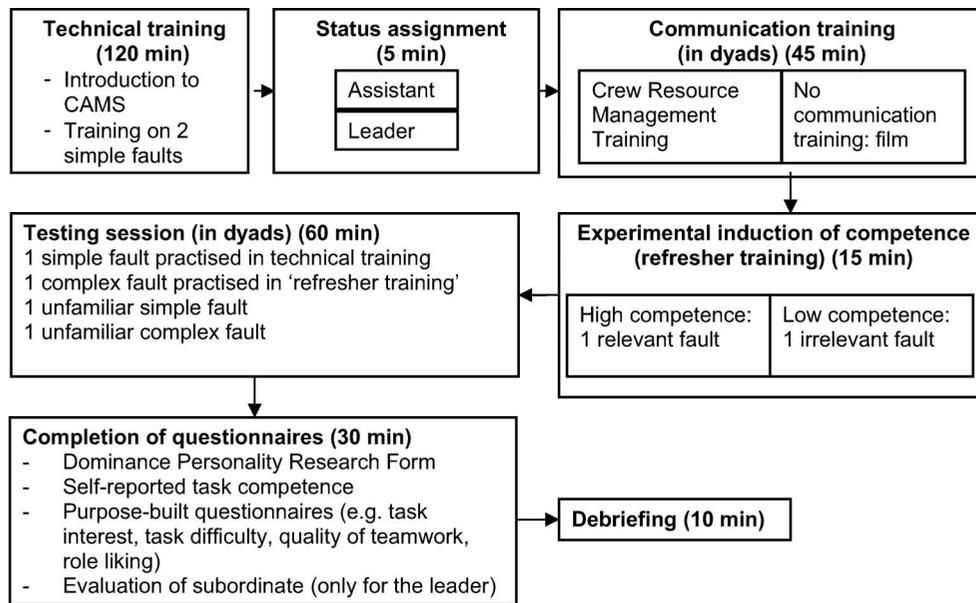


Figure 2. Experimental procedure (totalling about 7 h including breaks). CAMS = Cabin Air Management System.

themselves with the CAMS environment so that they were able to carry out all four tasks (system stabilisation, fault diagnosis, alarm acknowledgement and tank level recording) and to operate the technical system in a fault-free mode. In the second phase (also lasting about 60 min), participants were trained on two standard system faults ('leak in O₂ valve' and 'CO₂ scrubber ineffective'). The instructor went through a step-by-step procedure to teach the participant how each fault could be diagnosed and repaired (i.e. which symptoms correspond to each fault) and how the system state was stabilised most efficiently after the onset of the system fault. In a further training trial, the participants were asked to manage these two system faults independently, with the instructor being available for questions. To verify whether CAMS training was effective in equipping the trainees with the necessary skills, participants had to manage the leak in the O₂ valve by themselves, an activity in which all participants succeeded.

After the technical CAMS training, participants were given a light lunch during a 60-min break. After lunch, participants were paired into two-person teams, with TMs being randomly assigned to either the high or low status position (leader and assistant). Half of the teams were then randomly assigned to CST or the control condition.

2.4.2. Crew resource management training: communication skills training

CST (lasting about 45 min) started with a 4-min film scene from the National Geographic documentary

series 'Seconds from Disaster', entitled 'The crash of the century: Collisions on the runway'. This film shows a tragic accident in aviation, which was due to ineffective communication between the captain and his co-pilot. Participants watched the film and were asked: (a) to reflect on the attitudes of the captain and the co-pilot and the consequences; (b) to identify effective and ineffective communication patterns; (c) to suggest how communication patterns could be improved. Each participant was instructed to give his opinion according to the previous assigned status role (i.e. assistants answered questions relating to the co-pilot while leaders responded to the questions concerning the captain).

Participants then received further information about how communication can be more effective. The focus was put on two elements of communication, influencing skills and receptiveness.

In the final part of CST, participants took part in a role play in which they were already assigned the role they were going to adopt in the subsequent experiment (i.e. leader or assistant). The purpose of the role play was to give the participants actual experience in within-team communication to deal with difficult group processes. Participants received feedback from the experimenter about their performance in the role play. The training concluded with a short summary of the main points of CST.

The teams that did not receive CST watched and discussed a documentary instead ('Sur les routes d'Ushuaïa, Protéger les Paradis Terrestres: Retour à North Grip'). The documentary was about climate research in Greenland and completely unrelated to the

topic of communication skills. After watching the documentary, participants completed a questionnaire about its content (e.g. ‘what measures would you suggest to reduce CO₂ emission?’). The responses to the questions were then discussed in the group. This procedure ensured that the TMs without CST were equally familiar with each other as in the CST condition. This control condition was important because if an effect of CST on any of the outcome variables were to be found, this effect could not be explained by TMs just knowing each other better or having communicated with each other more but by the specific aspects of CST.

2.4.3. *Inducing task competence (‘refresher course’)*

Participants then received an additional training on CAMS (lasting 15 min) to create different levels of competence (high and low). This training was introduced to participants (with and without CST) as the ‘CAMS refresher course’. TMs were randomly assigned to either the high or low competence condition. The refresher course served the purpose of making specific TMs more competent and others less competent. Participants in the competent condition were trained on a complex fault (N₂ valve stuck open). Participants were told that this fault could not be repaired and that a control panel was blocked. Furthermore, it was pointed out that other control panels had to be used to stabilise the system. Finally, participants were given some practice on managing the system fault. This complex fault represented a different class of faults (i.e. it was not repairable and contained a concurrent failure of the most needed control panel). Being taught how to manage this fault took participants to a new level of system understanding. Participants in the low competence condition were trained on another standard fault (leak in mixer valve), which did not help them manage more efficiently the faults they were going to encounter in the testing session.

2.4.4. *Testing session and completion of questionnaires*

The testing session started with reinforcing the status roles allocated to the TMs (leader and assistant). The leader was given the ultimate decision-making responsibility and was told that he would evaluate the assistant’s performance after the testing session. The assistant was told to follow the leader’s instructions and support him. All of these status manipulations have been successfully used in other research (Galinsky *et al.* 2003, Schmid Mast and Hall 2003, Schmid Mast *et al.* 2008). The TMs’ respective roles were reinforced by various status symbols. Both were given badges indicating their roles. The leader was seated on a large

comfortable chair whereas the assistant was given a rather basic plastic chair, another power manipulation that has previously been used successfully (Chen *et al.* 2001). Furthermore, the leader was given control over the mouse and the keyboard, again to emphasise the status difference by giving higher control to the leader (a concurrent control of the system by both TMs was not possible due to constraints of the simulation environment).

Each team worked on a separate PC (three to six teams per session, supervised by up to three experimenters), with teams being placed at separate tables at a distance of approximately 2 m from each other. They were instructed to speak softly to ensure that within-team communication would not disturb other teams. The 60-min testing session consisted of a series of four fault scenarios that were presented in the following order: leak in O₂ valve (standard fault); N₂ valve stuck open (complex fault); dehumidifier set point failure (standard fault); cooler failure (complex fault). The faults lasted between 11 and 20 min, depending on the type of fault. At the end of each fault state, the system was reset (i.e. all parameters were stabilised) to avoid after-effects of previous unsuccessfully managed fault states). Between two fault states, fault-free phases were set up to include periods of low workload (lasting about 1 min).

After the testing session, participants were asked to fill in questionnaires that were used for manipulation checks and contained self-ratings of teamwork (a description of each instrument is given below). The order in which these questionnaires were presented is shown in Figure 2. Finally, they were debriefed and thanked for their participation.

2.5. *Measures*

2.5.1. *Performance measures*

Four performance measures were collected by the CAMS environment. The first two were defined as primary tasks, the last two as secondary tasks (with participants being instructed about the difference in their priorities).

- (1) System stabilisation errors: Sum of the duration (s), in which the control parameters are out of their target range.
- (2) Diagnostic errors: Number of incorrect diagnoses per fault state.
- (3) Alarm reaction time: Time elapsed (s) until a system alarm was acknowledged by clicking on a warning signal.
- (4) Prospective memory errors: Percentage of O₂ tank level recordings being missed.

Furthermore, information sampling behaviour (how many times the flow rates were checked per min) and system management strategies (number of attempts to repair a system fault that is not repairable) were measured.

2.5.2. Subjective state measures

Three subjective state measures were taken. Self-rated mental effort expenditure was measured with one item on a scale ranging from 0 (not at all) to 100 (a great deal) (mean 45.89, SD 23.72). Self-rated anxiety was assessed with one item on a scale ranging from 0 (not at all) to 100 (a great deal) (mean 31.38, SD 24.80). Finally, self-rated fatigue was measured with one item on a scale ranging from 0 (not at all) to 100 (a great deal) (mean 41.05, SD 23.70). The three subjective state measures were derived from the NASA-TLX (Hart and Staveland 1988) and have already been employed in a considerable number of studies (e.g. Hockey *et al.* 1998, 2007).

2.5.3. Questionnaire scales

2.5.3.1. Self-rated team performance. Self-rated team performance was measured with four items (two reverse scored) on a 5-point Likert format scale ranging from 1 (I strongly disagree) to 5 (I strongly agree). A sample item was: 'As a team we showed good task performance'. Items were averaged and larger number indicates higher self-rated team performance (mean 4.48, SD 0.52, Cronbach's $\alpha = 0.79$). The scale was purpose-built for this study.

2.5.3.2. Self-rated team involvement. Self-rated team involvement was measured with four items (two reverse scored) on a 5-point Likert format scale ranging from 1 (I strongly disagree) to 5 (I strongly agree). A sample item was: 'I did not put much effort into the task'. Items were averaged and larger number indicates higher self-rated team involvement (mean 4.52, SD 0.45, Cronbach's $\alpha = 0.73$). The scale was also purpose-built for this study.

2.6. Manipulation checks

2.6.1. Assigned status roles

To check whether the person in the assigned high status role really had more influence in the within-team interaction than the person in the assigned low status role, the self-reported influence that each TM had on the decision of the group was measured with one item ('How do you judge your influence on the group decision in comparison with your team mate?') on a scale from 1 (team mate's influence was much bigger) to 5 (my influence was much bigger). This scale was

developed for a previous study, which addressed the issue of team performance (Sauer *et al.* 2006).

To check whether both roles assigned were equally attractive to participants, self-reported role liking was measured with four items (two reverse scored). Sample items are 'I liked my role' or 'I had difficulty in accepting my role' (reverse scored). The four items were averaged (mean 3.98, SD = 0.89, Cronbach's $\alpha = 0.88$). As predicted, it was found that the person in the role of the leader (mean 3.81) felt more influential in decisions of the team than the person in the assistant's role (mean 3.13), $t(31) = 3.57$, $p = 0.001$. Also, as expected, participants did not differ in how much they liked their assigned roles (mean_{leader} 4.04, mean_{assistant} 3.92), $t(31) = 0.48$, $p = 0.64$. This scale was developed for a previous study on leader incompetence (Darioly and Schmid Mast 2010).

2.6.2. Task competence

To check whether the people who were trained on the complex fault (the competent ones) felt more competent during the interaction than the people who were not trained (the less competent ones), felt competence was assessed with one item ('How do you judge your knowledge about the simulation in comparison with your team mate?') on a scale from 1 (my team mate's knowledge was much bigger) to 5 (my knowledge was much bigger). It was predicted and found that the more competent participants (mean 3.19) reported to have more knowledge about CAMS than the less competent participants (mean 2.94), $t(62) = 2.21$, $p = 0.031$. This scale was developed for a previous study on team performance (Sauer *et al.* 2006).

2.6.3. Team collaboration

To check whether the CST teams perceived their collaboration to be better than teams without CST, each participant judged the quality of the collaboration with one item ('Our collaboration was good') on a scale from 1 (I strongly disagree) to 5 (I strongly agree). TM ratings were averaged. It was predicted and found that the teams who participated in CST perceived their team to collaborate better (mean 4.97) than teams who did not participate in CST (mean 4.63), $t(30) = 2.02$, $p = 0.05$. Again, this scale originates from a previous study on team performance (Sauer *et al.* 2006).

3. Results

To test the hypotheses about performance using team-based data, a two (with or without CST) by

two (congruent vs. incongruent teams) ANOVA was computed for each of the performance variables separately. For the analysis of questionnaire data of individual TMs, the same ANOVA was performed as for the performance measures but status was added as a third factor (leader vs. assistant) to the ANOVA model, with status being a repeated measures factor.

3.1. Cabin Air Management System performance measures

3.1.1. System stabilisation errors

An analysis of the time during which system parameters were out of their target range showed no effects between experimental conditions ($\text{mean}_{\text{grand}} = 176$ s). There was no significant main effect for congruence, $F(1,28) = 0.47$, $p = 0.501$, and none for CST, $F(1,28) = 0.87$, $p = 0.359$. There was no significant interaction, $F(1,28) = 0.86$, $p = 0.361$.

3.1.2. Diagnostic errors

With regard to the number of incorrect fault diagnoses per system fault ($\text{mean}_{\text{grand}} = 0.49$), there was no significant main effect for congruence, $F(1,28) = 0.67$, $p = 0.421$, and none for CST, $F(1,28) = 0.00$, $p = 1.00$. The interaction effect was also not significant, $F(1,28) = 2.67$, $p = 0.114$.

3.1.3. Alarm reaction time

In terms of time elapsed until a system alarm was acknowledged, there was no significant main effect for congruence ($\text{mean}_{\text{grand}} = 1.83$ s), $F(1,28) = 1.78$, $p = 0.193$, and none for CST, $F(1,28) = 0.84$, $p = 0.367$. The interaction effect was also not significant, $F(1,28) = 0.17$, $p = 0.680$.

3.1.4. Prospective memory errors

The analysis revealed that congruent teams (mean 3.63%) missed fewer tank level recordings than incongruent teams (mean 11.19%), $F(1,28) = 13.02$, $p = 0.001$, effect size $r = 0.56$. There was no significant main effect for CST, $F(1,28) = 0.62$, $p = 0.439$, but a significant interaction effect, $F(1,28) = 5.55$, $p = 0.026$, effect size $r = 0.56$, for which the means of each condition are shown in Figure 3. The results indicate that the incongruent teams benefited from CST whereas the congruent teams did not.

3.1.5. Information sampling

Examining the number of flow meter viewings showed that incongruent teams (mean 5.44) sampled the flow

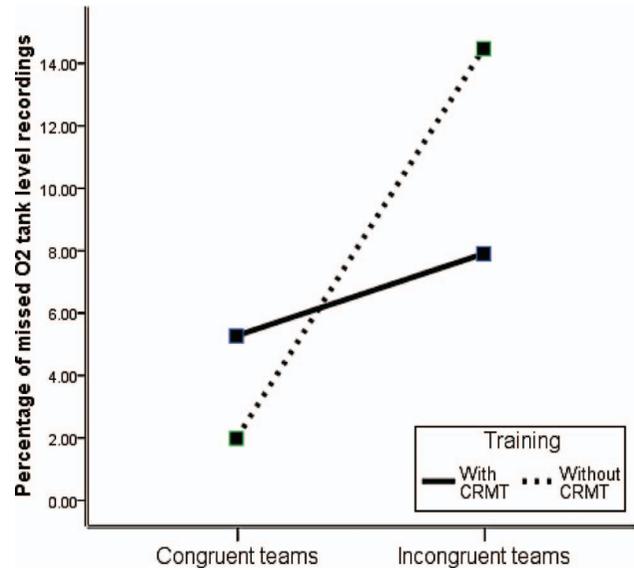


Figure 3. Prospective memory failures as a function of crew resource management training (CRMT) and team congruence.

rates more frequently than the congruent teams (mean 2.09), $F(1,28) = 9.31$, $p = 0.005$, effect size $r = 0.50$. There was no significant main effect for CST, $F(1,28) = 0.45$, $p = 0.507$, and no significant interaction effect, $F(1,28) = 0.04$, $p = 0.846$.

3.2. Cabin Air Management System subjective state measures

The same ANOVA was performed as for the performance measures. However, status was added as a third factor (high or low status) to the ANOVA model, with status being a repeated measures factor.

3.2.1. Mental effort

The results showed that TMs in high power positions reported having expended more effort into the task activities (mean 50.00) than TMs in low power positions (mean 41.78), $F(1,28) = 4.22$, $p = 0.049$. There was no significant main effect for congruence, $F(1,28) = 1.34$, $p = 0.256$, and none for CST, $F(1,28) = 0.19$, $p = 0.664$. None of the interaction effects was significant (all F 's < 1.39 , all p 's > 0.248).

3.2.2. Anxiety

The results showed that incongruent teams (mean 39.41) were more anxious than congruent teams (mean 23.34) $F(1,28) = 6.37$, $p = 0.018$. There was no significant main effect for CST, $F(1,28) = 0.16$,

$p = 0.69$, and no significant main effect of status, $F(1,28) = 2.42$, $p = 0.131$. None of the interaction effects was significant (all F 's < 1.41, all p 's > 0.245).

3.2.3. Fatigue

TMs in low status positions felt more fatigued (mean 47.44) than TMs in high status positions (mean 34.66) $F(1,28) = 4.92$, $p = 0.035$. There was no significant main effect for congruence, $F(1,28) = 0.29$, $p = 0.592$, and none for CST, $F(1,28) = 0.60$, $p = 0.446$. None of the interaction effects was significant (all F 's < 2.06, all p 's > 0.162).

3.3. Questionnaire scales

3.3.1. Self-rated team performance

Teams with CST rated their performance higher (mean 4.69) than teams without CST (mean 4.26), $F(1,28) = 7.87$, $p = 0.009$, effect size $r = 0.47$. There was no significant main effect for congruence, $F(1,28) = 0.11$, $p = 0.746$, and none for status, $F(1,28) = 0.71$, $p = 0.407$. None of the interaction effects was significant (all F 's < 1.39, all p 's > 0.248).

3.3.2. Self-rated team involvement

Teams with CST rated their team involvement marginally significantly higher (mean 4.59) than teams without CST (mean 4.38), $F(1,28) = 3.46$, $p = 0.074$, effect size $r = 0.33$. There was no significant main effect for congruence, $F(1,28) = 0.49$, $p = 0.491$, and none for status, $F(1,28) = 0.93$, $p = 0.344$. None of the

interaction effects was significant (all F 's < 1.63, all p 's > 0.212).

4. Discussion

The first goal of the present study was to meet the demands of Salas *et al.* (2001, 2006) that multi-level analyses should be employed to evaluate the effectiveness of CRM training. There are some implications from this laboratory-based study with regard to the methodological approach that may be used for a field-based evaluation of CRM training.

The second goal of the present study was to investigate how teams in which the status hierarchy either matched the task-competence hierarchy or not performed on different performance levels and whether CST attenuated the expected negative effect of mismatched teams. The results showed that, on a few tasks, performance was better for congruent in comparison with incongruent teams and for one task CST was beneficial specifically for incongruent teams. Therefore, the hypotheses were only partially supported. In other words, when effects were found, they went in the predicted direction but it also became clear that only a few effects were found.

The findings of the study are presented, before their implications for the methodological approach are considered. A summary of the findings is provided in Table 1.

4.1. Effects of team congruence

It was predicted that, in incongruent teams, there would be more negative outcomes (notably decreased

Table 1. Effects of communication skills training (CST), team congruence and status hierarchy.

Variable	CST effect	Congruence effect	Status effect
CAMS performance measures			
System stabilisation errors	=	=	NA
Diagnostic errors	=	=	NA
Alarm reaction time	=	=	NA
Prospective memory errors*	=	congr < incongr	NA
Information sampling	=	congr < incongr	NA
System management	with > without	=	NA
CAMS subjective state measures			
Mental effort	=	=	high > low
Anxiety	=	congr < incongr	=
Fatigue	=	=	high < low
Questionnaire			
Self-rated team performance	with > without	=	=
Self-rated team involvement	with > without	=	=

CAMS = Cabin Air Management System; congr = congruent teams; incongr = incongruent teams.

Note: NA = not applicable (because there were only team level data and measures were not separated for leaders and assistants).

*Indicates the only variable with an interaction effect, which is depicted in Figure 2. = Indicates no difference between the groups: with = with CST; without = without CST.

performance) than in congruent teams. Although the non-optimal team composition did not lead to dramatic performance decrements, it still affected some performance aspects, especially for non-optimal teams without CST. This suggests that even in the absence of CST, teams were well able to maintain primary task performance under non-optimal team composition. However, this maintenance of primary task performance may come at a cost, as suggested by the model of compensatory control (Hockey 1997). This was observed in the form of decrements in secondary task performance on the prospective memory task due to a shift in resources from secondary tasks to primary tasks to maintain overall performance levels.

In addition to the secondary task performance impairment, further evidence for the problems ensuing from an ill-matching team composition also stemmed from flow meter sampling behaviour. Incongruent teams sampled the flow meter more often than necessary, which may be due to the greater ambiguity of their roles. Since the incongruent teams were composed of a less competent leader and a more competent assistant, it is not known whether the performance decrement was due to the less competent leader or to the more competent assistant or to both. Expectation States Theory (Berger *et al.* 1977) suggests that the leader is in a high status position that may be considered illegitimate. This may have put strain on him, resulting in performance decrements. However, the illegitimacy might also make the leader want to compensate for his lack of competence by increasing effort in simple activities (e.g. flow meter reading). On the other hand, it is also possible that the competent assistant is responsible for the results found. He might have felt reluctant to demonstrate his competence, which he might have perceived as overstepping his role. This might have entailed the performance decrement and he might have diverted his efforts to a relatively insignificant activity, such as checking the flow meters. Future research needs to address this point and disentangle completely the competence and status positions by also examining teams with competent leaders and competent assistants as well as teams with low-competence leaders and low-competence assistants.

For the variables that were assessed for leaders and assistants separately, it was possible to look at whether there were status differences. It was found that the team leaders reported to having expended more mental effort into the task than assistants and that assistants felt more fatigued than leaders. These findings are in line with the high status and low status roles, with the leader being ultimately responsible for task completion. Therefore, there is more at stake for the leader,

which may explain his higher expenditure of mental effort. The increased fatigue experienced by assistants was also not surprising because his role was clearly less interesting. The manipulation of the mouse, for instance, was a responsibility assigned to the leader, which may have caused the assistant to feel less involved at times. Previous research has also indicated that assigning a (even minor) functional role to one TM (e.g. being in charge of making system interventions by mouse and keyboard) has resulted in increases in subjectively experienced operator strain, which became evident in measures such as effort, anxiety and fatigue (Sauer *et al.* 2006).

4.2. Effects of communication skills training

The lack of strong evidence for the benefits of receptiveness and influencing skills may be due to the general difficulties of demonstrating the effectiveness of training methods at the behavioural level while evidence at the levels of reaction and learning was easier to observe. Overall, the present study seems to confirm the overall pattern that emerged from the two literature reviews (Salas *et al.* 2001, 2006), showing that the further one moves down the levels of Kirkpatrick's classification scheme, the less strong the evidence for the effectiveness of receptiveness and influencing skills becomes. This became evident at the first level (subjective reactions to training), which showed that self-rated team involvement was perceived to be more positive in teams with CST as compared with teams without CST. At the second level (learning), benefits were also found for CST, with CST teams considering themselves as better performers than the control group. At the third level (behaviour), it emerged that primary task performance did not benefit from CST. However, CST buffered the performance decrements on secondary tasks that were recorded for unfavourable team compositions (i.e. incongruent teams). Incongruent teams with CST were better at recording O₂ tank levels throughout the testing session than those without CST. This effect was not found for congruent teams. This finding may be interpreted within Hockey's (1997) model of compensatory control. It assumes that under unfavourable working conditions (here, incongruent teams), performance decrements may be expected but these decrements are rarely observed on primary tasks but are more likely to emerge in secondary tasks, since these are more sensitive to variations in workload. Since receptiveness and influencing skills provided the basis for more effective within-team communication in the incongruent teams, this allowed for more cognitive resources to be committed to task completion with the incongruent CST teams (compared with incongruent

teams without CST). This led to improvements in performance for the prospective memory task, which is characterised by considerable demands on working memory. CST appears to be most beneficial in situations for which it was originally designed, such as sub-optimal team composition. The non-task skills taught by CST were of greater benefit under these latter circumstances. Future research may wish to examine whether this observation extends to other sub-optimal working conditions that impinge on the quality of team collaboration in a more indirect way (e.g. time pressure, noise). These sub-optimal working conditions also put strain on the team, which might then be alleviated by CST because it allows for a more flexible adaptation of within-team system management strategies (e.g. high noise levels may require the shift to a cognitively less demanding system management strategy). A further step would be to identify the most important elements of CST and to focus on these and possibly place less emphasis on elements that were not found to be strongly related to performance.

4.3. Multi-level analyses and implications for field-based evaluation of crew resource management training

Against the background of the findings of the present study, the points made by Salas *et al.* (2001, 2006) can only be reiterated. There is a need for a multi-level analysis (in the sense of Kirkpatrick's (1976) classification scheme) to determine the multiple outcomes of a given CRM training. In practical terms, this may mean that, in addition to subjective reactions as the first level (which is typically measured in CRM training evaluation), measures need to be taken, at least, at one other level, which should preferably be behaviour at the third level. High fidelity simulators may be employed for that purpose since they provide a very safe environment for measuring performance.

The four levels in Kirkpatrick's model may be considered to represent a vertical dimension of the classification scheme. However, it may not be sufficient in CRM training to consider the vertical dimension only (i.e. covering as many levels of Kirkpatrick's model as possible), one may also need to take into account the horizontal dimension (i.e. using several parameters to measure the effects of CRM training at the same level). Although the concerns of Salas and his colleagues have mainly focused on the vertical dimension, the horizontal dimension has also been addressed implicitly. The significance of the horizontal dimension may need to be emphasised more strongly. Hockey's (1997) model may be useful in this respect, as it focuses on the horizontal dimension by emphasising the importance of measuring different aspects of performance (i.e. behaviour in Kirkpatrick's terms). For the

horizontal dimension, this may mean in practical terms that, at the behavioural level, at least one secondary task should be included. This secondary task needs to be relevant to the technical environment to which the CRM training is applied. The more resource-sensitive secondary tasks are important, as they allow one to gain a better understanding of the overall demands of the work environment than primary tasks do (Wickens and Hollands 2000). The idea of using several measures along the horizontal dimension also applies to other levels of Kirkpatrick's classification scheme. At the reaction level, for example, different aspects of subjective reactions may be measured (e.g. team climate, TM involvement).

The criticism expressed by Salas *et al.* (2001, 2006) also referred to the lack of experimental research in the field of CRM training evaluation. While it is acknowledged that experimental studies are difficult to conduct in the field, it would make it possible to look more closely at specific elements of CRM training and identify those that seem to be critical for team performance. For example, apart from communication skills, other elements of CRM training may also be examined with regard to their impact (e.g. human factors knowledge). A combination of laboratory-based as well as field-based experimental work would be most promising, making use of the advantages of rigorous testing in the laboratory combined with the high ecological validity offered by the use of high-fidelity simulators.

4.4. Methodological issues and limitations

The study entails a number of limitations, which are largely owed to the constraints of carrying out laboratory-based experimental research. First, the current sample has a lower level of technical expertise than recipients of real CRM training in the field, who tend to be subject matter experts. This represents an inherent problem of using an experimental approach that aims to manipulate task competence. Novices were selected as participants and appropriately trained because the task competence manipulation would not have been feasible with subject matter experts. Since the level of task competence used in a study is not only a function of training time but also a function of system complexity (which was much lower in the present case than for technical systems typically used for CRM training), the present experimental set-up is considered to represent an acceptable laboratory-based simulation of a real work environment.

Second, the time allocated to CST was rather short, again owing to budgetary and time constraints of the study. The short training time may have led to the overall rather minor differences between experimental

conditions. One may speculate that more extensive CST might have resulted in more substantial effects. The reduced time allocated to CST also required focusing on selected elements of CRM training rather than modelling the full CRM training programme. Therefore, two elements of CRM training (receptiveness and influencing skills) were employed in the present study. Although there are no independent means to test the degree to which these two elements are related to performance, it is believed that due to the complex nature of the task environment, incongruent teams need to communicate to make best use of their team resources. Without communication, the low-competence leader would need to complete the task alone since the competent assistant would have no means of contributing to team performance, as he did not have access to any of the controls. Therefore, the present authors strongly endorse that receptiveness and influencing skills were of benefit in the present study since both improved the quality of within-team communication. Please note that in the present study, it is not known whether it was more the receptiveness or the influencing skill aspect of CST that drove the effect.

Third, one can question the overall ecological validity of the laboratory-based experiment. However, the advantage of using an experimental approach (albeit lacking ecological validity by definition) is that it makes it possible to test causal influences. Given that much of the existing literature in this field is correlational, the present authors think that there is some merit in employing an experimental approach.

4.5. Conclusion

Although no strong evidence for the benefits of training receptiveness and influencing skills could be demonstrated in the present study, the work has some merit by providing a good example of the methodological principles that should guide an evaluation of CRM training. This primarily refers to the multi-level analysis of the multiple outcomes of CRM training (i.e. along the vertical dimension) but also includes the use of multiple measures at the same level (i.e. horizontal dimension). Furthermore, it demonstrates that an experimental approach can be employed to examine the effects of CRM training in a laboratory setting before the transferability of CRM training to a work setting is tested.

References

- Bass, B.M., 1990. *Bass & Stogdill's handbook of leadership: Theory, research, and managerial applications*, 3rd ed. New York: Free Press.
- Berger, J., et al., 1977. *Status characteristics and social interaction: An expectation states approach*. New York: Elsevier.
- Chen, S., Lee-Chai, A.Y., and Bargh, J.A., 2001. Relationship orientation as a moderator of the effects of social power. *Journal of Personality & Social Psychology*, 80 (2), 173–187.
- Darioly, A. and Schmid Mast, M., 2010. Facing an incompetent leader: the effects of a non-expert leader on subordinates' perception and behavior. *European Journal of Work and Organizational Psychology*. Published on iFirst 14 July 2010 DOI: 10.1080/13594320903429576. <http://dx.doi.org/10.1080/13594320903429576>
- Droog, A., 2004. The current status of CRM training and its regulations in Europe. In: K.M. Goeters, ed., *Aviation psychology: Practice and research*. Aldershot: Ashgate Publishing Limited, 221–230.
- Dunlap, J.H. and Mangold, S.J., 1998. *Leadership/followership recurrent training: student manual (AAR-100)*. Washington, DC: Federal Aviation Administration.
- Falk, I., 1999. Situated leadership: A new community leadership model. In: Rural community partnerships: Education for the 21st century. *Proceedings of the National Rural Education Conference of the Country Education Project*, held at School of Mines, Ballarat 1–3 October 1998. Country Education Project. Available at: <http://crlra.utas.edu.au/pages/files/discussion/1998/D02-1998.pdf>
- Galinsky, A.D., Gruenfeld, D.H., and Magee, J.C., 2003. From power to action. *Journal of Personality and Social Psychology*, 85 (3), 453–466.
- Ginnett, R.C., 1993. Crews as groups: their formation and their leadership. In: E.L. Wiener, B.G. Kanki, and R.L. Helmreich, eds., *Cockpit resource management*. Orlando, FL: Academic Press, 71–98.
- Grote, G., et al., 2010. Adaptive coordination and heedfulness make better cockpit crews. *Ergonomics*, 53 (2), 211–228.
- Hamblin, R.L., Miller, K., and Wiggins, J.A., 1961. Group morale and competence of the leader. *Sociometry*, 24 (3), 295–311.
- Hart, S.G. and Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: P.A. Hancock and N. Meshkati, eds. *Human mental workload*. Amsterdam: North-Holland, 139–183.
- Hockey, G.R.J., 1997. Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. *Biological Psychology*, 45 (1–3), 73–93.
- Hockey, G.R.J., Sauer, J., and Wastell, D.G., 2007. Adaptability of training in simulated process control: comparison of knowledge- and rule-based guidance under task changes and environmental stress. *Human Factors*, 49 (1), 158–174.
- Hockey, G.R.J., Wastell, D., and Sauer, J., 1998. Effects of sleep deprivation and user interface on complex performance: a multilevel analysis of compensatory control. *Human Factors*, 40, 233–253.
- Katz, D. and Kahn, R., 1952. Some recent findings in human relations research in industry. In: G. Swanson, T. Newcomb, and E. Hartley, eds., *Readings in social psychology*. New York: Holt, 295–311.
- Kirkpatrick, D.L., 1976. Evaluation of training. In: R.L. Craig, ed., *Training and development handbook: a guide to human resource development*. New York: McGraw-Hill, 301–319.

- Littlepage, G.E., 1991. Effects of group size and task characteristics on group performance: a test of Steiner's model. *Personality and Social Psychology Bulletin*, 17 (4), 449–456.
- Pearce, C.L. and Sims, H.P. Jr, 2002. Vertical versus shared leadership as predictors of the effectiveness of change management teams: an examination of aversive, directive, transactional, transformational, and empowering leader behaviors. *Group Dynamics: Theory, Research, & Practice*, 6, 172–197.
- Perry, M.L., Pearce, C.L., and Sims, H.P. Jr., 1999. Empowered selling teams: how shared leadership can contribute to selling team outcomes. *Journal of Personal Selling & Sales Management*, 19 (3), 35–51.
- Salam, S., 1998. The effects of subordinate competence, leader competence, leader integrity, and technology on subordinate participation seeking, performance, satisfaction, and agreement. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 59.
- Salas, E., Rhodenizer, L., and Bowers, C.A., 2000. The design and delivery of crew resource management training: exploiting available resources. *Human Factors*, 42 (3), 490–511.
- Salas, E., *et al.*, 2001. Team training in the skies: does crew resource management (CRM) training work? *Human Factors*, 43 (4), 641–674.
- Salas, E., *et al.*, 2006. Does crew resource management training work? An update, an extension, and some critical needs. *Human Factors*, 48 (2), 392–412.
- Sauer, J., Wastell, D.G., and Hockey, G.R.J., 2000. A conceptual framework for designing micro-worlds for complex work domains: a case study of the Cabin Air Management System. *Computers in Human Behavior*, 16 (1), 45–58.
- Sauer, J., *et al.*, 2003. Performance in a complex multiple-task environment: effects of occasional night work and extended task involvement. *Human Factors*, 45 (4), 657–669.
- Sauer, J., *et al.*, 2006. Cognitive diversity and team performance in a complex multiple task environment. *Ergonomics*, 49 (10), 934–954.
- Schmid Mast, M. and Hall, J.A., 2003. Anybody can be a boss but only certain people make good subordinates: behavioral impacts of striving for dominance and dominance aversion. *Journal of Personality*, 71 (5), 871–891.
- Schmid Mast, M., Jonas, K., and Hall, J.A., 2008. Give a person power and he/she will show interpersonal sensitivity: the phenomenon and its why and when. *Journal of Personality and Social Psychology*, 97 (5), 851–865.
- Steiner, I.D., 1972. *Group processes and productivity*. New York: Academic Press.
- Wickens, C.D. and Hollands, J.G., 2000. *Engineering psychology and human performance*. Upper Saddle River, NJ: Prentice Hall.