



Universiteit Leiden

Psychologie
Faculteit der Sociale Wetenschappen



Flightcrew Crew Resource Management training

An evaluation and effectiveness study

Name: Michiel de Galan
Student number: 0721875
Supervisor: Guido Band
Second reader: Fenna Poletiek
Cognitive Psychology
Thesis MSc Applied Cognitive Psychology

Abstract

The research question for this study was twofold. First, does KLM Royal Dutch Airlines' three-day classroom Crew Resource Management (CRM) training have a positive influence on flightcrews' knowledge, skills, and attitudes? Second, does the introduction of Behavior-Modeling Training (BMT), i.e. making training goals explicit, modeling desired behavior, increasing training opportunities, providing actionable feedback related to goals, and maximizing transfer, lead to improved training outcomes when compared to non-altered courses? 68 participants in KLM's Crew Management Courses (CMC) were assessed. Their verbal and metacognitive knowledge, self-efficacy, and attitudes and attitude strength were measured pre- and post-training with a questionnaire that was developed specifically for this purpose. Reactions to the training were measured post-training only and were significantly on the positive side of neutral. Overall, the effect of CMC training on participants' knowledge, self-efficacy, and attitudes was positive and significant. The largest learning effects were found for knowledge and skill, consistent with prior research. The effect on attitude was slightly smaller, which could be due to the fact that CRM has a long tradition, and is well accepted within KLM.

BMT participants did not show significantly more improvement than non-BMT participants, although metacognitive knowledge and self-efficacy tended to improve more, and reactions were slightly more positive in the BMT group, just short of significance. Furthermore, participants improved most on items that were especially relevant for the effects of BMT, and instructors and candidates report that they like the way of training and the structure it provides. Thus, although no definitive conclusions about the effectiveness of BMT can be drawn, the results suggest that it is a promising candidate for CRM training improvement.

Keywords: crew resource management, behavior-modeling training, effectiveness, evaluation

Crew Resource Management, an Evaluation and Effectiveness Study

In the morning on November 4th, 2010, a Qantas Airbus A380 departed from Singapore. About 4 minutes after takeoff, two ‘bangs’ were heard and a multitude of warnings emerged on the electronic centralized aircraft monitor. The aircraft had sustained significant damage from debris due to the explosion of the number 2 engine. Its structure and systems were damaged, and an uncontained engine fire occurred. Nevertheless, all 469 people on board returned safely to the ground; not in the least due to the impeccable teamwork of the five pilots. They managed to determine how much of the aircraft was still functioning and eventually landed an overweight aircraft, on the edge of a stall with marginal control effectiveness. On the ground, an engine would not shut down, preventing evacuation from an aircraft that was still on fire. Eventually the engine was put out with firefighting foam and everyone could leave the aircraft safely. Qantas flight 32 has become known as a prime example of successful Crew Resource Management (CRM); the crew’s outstanding airmanship, workload management, problem solving, stress management, and teamwork were main factors in determining a positive outcome.

What is CRM?

CRM can be defined as “the effective use of all available resources, i.e. equipment, procedures and people, to achieve safe and efficient flight operations” (Civil Aviation Authority, 2002). It is concerned with the cognitive and interpersonal skills needed to manage the flight within an organized aviation system and encompasses a wide range of knowledge, skills and attitudes. With the introduction of reliable multi-pilot turbojet aircraft in the 1950s, safety research indicated that human error had become the main causal factor in aviation accidents (Helmreich, & Clayton Foushee, 2010; O'Connor, Hörmann, Flin, & Lodge, 2002). This led the industry to implement training programs aimed at crew coordination and flight deck management. First, they were based on managerial training programs from the corporate

domain. Quickly however, courses began to focus on group dynamics and the system in which crews must function. Then, around 1990, CRM elements became integrated into technical training and procedures (Civil Aviation Authority, 2002; Helmreich, Merritt, & Wilhelm, 1999).

Typical introductory CRM training is conducted in a classroom for 2 or 3 days and is required by aviation regulation (EASA, 2014; FAA, 2016). Teaching methods are lectures, practical exercises, role-playing, case studies and videos. Refresher training is typically classroom-based and conducted yearly, in a half-day or whole-day course. These refresher trainings are sometimes integrated into (partly) simulator-based trainings. Furthermore, whenever crewmembers change aircraft, operator, or crew position, CRM elements are integrated into the course.

CRM at KLM

Pilots at KLM receive fairly typical CRM training, consisting of introductory classroom training, refresher trainings, and conversion and command course trainings. The introductory training is a 3-day classroom training, called Crew Management Course 1 (CMC-1). It is delivered by trained, in-house CRM instructors and covers: human performance and limitations, threat and error management, personal awareness, stress and stress management, fatigue and vigilance, assertiveness, situational awareness, automation, communication, leadership and coordination, resilience development, surprise and startle, cultural differences, safety culture, and case studies.

CRM refresher trainings are part of the technical trainings that are held twice yearly, Type Recurrents (TRs). These TRs consist of a two-hour briefing session followed by a three-and-a-half-hour simulator session. Approximately half an hour briefing time in each TR is devoted to CRM topics, which are then practiced in the simulator. Additionally, a one-day flight safety refresher training is attended every three years, of which approximately two

hours are devoted to CRM training. CRM elements are integrated in the conversion and command courses. Additionally, the command course training comprises a stand-alone, classroom training, Crew Management Course 2 (CMC-2). The same elements as in CMC-1 are covered, but with a more personalized approach and a focus on leadership.

Does CRM training work?

Several frameworks for training evaluation have been proposed, of which Kirkpatrick's (1976) 4-level model is still the most widely used (Alvarez, Salas, & Garofano, 2004; Passmore & Velez, 2014; Shuffler, Salas, & Xavier, 2010). Kirkpatrick (1976) identified: reactions, learning, behavior, and results. *Reactions* measures trainee satisfaction, *learning* the increase of knowledge in trainees, *behavior* the degree to which trainees apply what they learned on the job, and *results* the eventual effect on the organization.

Salas, Burke, Bowers, and Wilson (2001) reviewed 58 studies of aviation CRM training and found that affective and utility reactions were positive. On the learning level, positive changes in attitudes and knowledge were found. Behavioral-level results were more mixed, but most studies found a positive effect on behavior. The impact on the 4th level, results, remained inconclusive; the few studies that reported bottom-line evidence, lacked important details and information. Overall, their evidence suggests that CRM training is effective, but they conclude that multilevel evidence is scarce.

Salas, Wilson, Burke, and Wightman (2006) conducted another review of studies that were published after the initial review, as well as studies conducted in fields outside aviation. About half of the 28 studies showed mixed results. Nevertheless, the impact on reactions and attitudes appeared robustly positive. For learning and behavior, about half of the studies found positive, and half of the studies mixed results. Only five studies evaluated results; three of them found positive results, and two mixed or no results. The authors conclude that there is evidence that CRM training is effective at some levels, but that the picture is still unclear.

A third review of CRM training effectiveness was performed by O'Connor and colleagues (2008). In their quantitative assessment of 16 published studies of CRM training they found large significant benefits for reactions and attitudes. For knowledge and behavior, the effects were medium and large, respectively, but not significant. Therefore, these results should be interpreted with caution, but they are positive and consistent with what other studies found.

In summary, the literature strongly suggests that CRM training is effective at some levels. However, Salas, et al (2001), Salas, et al (2006) and O'Connor et al (2008), all comment on the limited scale of studies and stress the importance of continued evaluation. Furthermore, because few studies assessed the specific effects of CRM training, there are few scientifically validated training programs, resulting in a lack of theoretical guidance on implementing effective programs (Civil Aviation Authority, 2003).

Aims of the study

This study will try to fill these voids by investigating the CMC 1 and 2 classroom-trainings at KLM. The goal is twofold: (1) begin to establish a continuing, systematic process of training evaluation at KLM, and (2) compare the effectiveness of two types of training, to provide theoretical guidance for the design of training programs.

Training evaluation. A CRM training-evaluation tool (questionnaire) was developed, which assessed level 1 and 2 of Kirkpatrick's hierarchy. The instrument had to be easy-to-use and concise, therefore level 3 (behavior) and 4 (results) were not included. CRM assessment capabilities are currently being integrated into KLM's behavior assessment tool that is used by instructors, making it possible to validate the questionnaire with observational measures later on, bridging the gap towards level 3 (behavior).

Increasing training effectiveness. A preliminary analysis of the CMC courses compared it to best practices and recommended that more effective exercises should be

provided (De Galan, 2016). In training aimed at skills, most learning occurs through deliberate practice and feedback (Salas, Tannenbaum, Kraiger, & Smith-Jentsch, 2012), and without practice, little learning occurs for CRM-related behaviors (Morin, 1998; Smith-Jentsch, Salas, & Baker, 1996). Instructors reported that they lacked support in conducting, and providing feedback on practical exercises, and that exercises were therefore performed less well than considered optimal (De Galan, 2016). This study assessed whether introducing Behavior Modeling Training (BMT) would increase training effectiveness.

Behavior Modeling Training. BMT is based on Bandura's (1977b) social learning theory. The theory emphasizes the social nature of learning, and the importance of role models and social reinforcement. It has five focal points, namely: (a) clearly define the behaviors to be learned, (b) model the use of those behaviors, (c) provide opportunities for active practice, (d) provide feedback, and (e) take steps to maximize transfer (Taylor, Russ-Eft, & Chan, 2005). Early investigations revealed positive and large training effects, but were based on a limited number of studies (M. J. Burke & Day, 1986; Smith-Jentsch et al., 1996). A recent meta-analysis of 117 studies, by Taylor, Russ-Eft and Chan (2005) included unpublished studies, and studies with negligible effects. Although they found more variation than earlier reviews (Arthur, Bennett, Edens, & Bell, 2003; M. J. Burke & Day, 1986), they still reported a large effect on knowledge and skill, and a medium effect on attitude. The effect was persistent, especially for skills, which were maintained or improved over time, but not for declarative knowledge, which decayed over time. The authors conclude that their results warrant the use of BMT in organizations.

The exercises in the CMC curriculum of KLM were adjusted according to BMT. For maximum effectiveness, suggestions from the science of training were used as well. For the BMT condition, the following was incorporated: (a) Provide clear goals. Installs a framework for self-directed learning (Hattie & Timperley, 2007), and stimulates self-regulation (Newell,

Lagnado, & Shanks, 2015). As self-regulation increases training effectiveness, goals should be set by participants themselves as much as possible (Salas et al., 2012). **(b) Modeling.** Has been shown to increase training effectiveness, this goes for non-effective models as well (Salas et al., 2012). **(c) Deliberate practice.** Most skill learning occurs this way (Salas et al., 2012). **(d) Feedback.** Its effects are strongly dependent on context, and type of feedback (Cook et al., 2013; Salas et al., 2012). Feedback should be clear, actionable, related to goals, and enable development of accurate mental models. **(e) Maximize transfer.** Making the connection between the content and the job (Taylor et al., 2005), and including challenging and variable exercises that resemble the transfer environment will benefit transfer (Salas & Rosen, 2010). *Error training* was also used: Trainees should learn to cope with errors both on a strategic and an emotional level. Error training is associated with better performance, and higher self-efficacy (Salas & Rosen, 2010). Scenarios were already sufficiently challenging, so that trainees would make errors; several exercises were rewritten to include ‘trying several ways to approach the problem’. The exercises were added to the BMT summary described in the procedure section, see Appendix A.

Research question & hypotheses

In summary, this study had two focal points. First, an evaluation of CRM training was performed, at the levels of *reactions*, and *learning*. A pre- and post-questionnaire was developed to measure CMC-1 and 2. Measurement after training was compared with measurement before; a follow-up measurement after 6 months is planned. The main outcomes of interest were three learning categories: knowledge (K), skill (S), and attitude (A). Additionally, trainee reactions were collected after training. Second, an experiment was conducted, comparing a BMT version, and an unaltered, control version of the CMC training. There were two research questions with two hypotheses:

1. Are KLM's CMC training courses achieving their intended learning outcome in terms of KSAs and reactions?

- Hypothesis 1 stated that KLM's CMC courses would improve participants' KSAs, and reactions.

2. Does the introduction of BMT improve learning and reactions, as compared to the unaltered training?

- Hypothesis 2 stated that the adapted CMC curriculum, based on BMT, would improve participants' KSAs and reactions more than the unaltered control condition.

Measurements

Outcome measures were based on Kraiger and colleagues (1993), who expanded on Kirkpatrick's (1976) second level, learning. They defined three major learning outcome categories: *cognitive*, *skill-based*, and *affective*. A questionnaire was developed to measure verbal (VK) and metacognitive knowledge (MK); self-efficacy (SE); and attitudes (AT) and attitude importance (AI). The resulting three main scales closely mirrored Kraiger's (1993) main categories, see Table 1. However, contrary to Kraiger's (1993) classification, in this study, SE was used not as an affective outcome, but to infer skill learning. Validity was assessed by an expert panel, and reliability was assessed by conducting a pilot among 45 KLM pilots. A description of the full process used to create and test the questionnaire can be found in De Galan (2017). Here the most important aspects are summarized.

Table 1.

Kraiger et al. (1993)	Cognitive outcomes	Skill-based outcomes	Affective outcomes
This study Main scales	Knowledge (K)	Skill (S)	Attitude (A)
Subscales	VK MK	SE	AT AI

Response-shift bias

Changes in raters' standards can influence results (Howard et al., 1979). This has special relevance for interventions that are supposed to change participants' understanding or awareness of the variable being measured, such as CRM training. Put simply, participants view the target variable differently after the intervention than before, leading a pretest-posttest comparison to compare different attributes. The issue has been termed *response-shift*, and is hypothesized to consist of three interrelated aspects: a change in internalized standard, a change in values or priorities, or a change in definition of the target construct (Schwartz & Sprangers, 1999; Schwartz, Sprangers, Carey, & Reed, 2004). It exerts its influence in many domains, such as healthcare (Schwartz & Sprangers, 1999), social interventions (Hill & Betz, 2005), teaching (Cartwright & Atwood, 2014), and many areas of training, including CRM training (Rosen et al., 2012; Sprangers & Hoogstraten, 1989). Response-shift was assessed for metacognitive knowledge, self-efficacy, and attitudes using the thentest approach (Howard et al., 1979): Participants reported how they perceived themselves at the present time (post), and immediately following each item, reported on the same item again, this time how they perceived themselves to have been just before training (then).

Reactions

Four reactions questions were included (posttest), assessing whether participants liked the training, whether it increased their ability to perform their job, whether they thought it was useful, and whether it would influence their behavior.

Knowledge

Metacognitive knowledge appeared most useful for CRM evaluation. It refers to knowledge of what strategies one has available, and how and when they are most effective (Ford, Kraiger, & Merritt, 2010). Ten questions were developed based on the Metacognitive Awareness Inventory (MAI) (Schraw & Dennison, 1994). Items assessed whether trainees

had been generating and testing hypotheses, were operating under goals, were aware of their mistakes, and determined their level of proceduralization. A 12-point Likert scale was used, labeled at five points with: 'not true at all', 'mostly untrue', 'neutral', 'mostly true', and 'completely true'. Possible scores ranged from 1 (not true at all) through 12 (completely true). Although not a primary goal of CRM training, five multiple-choice verbal knowledge questions were developed, assessing five prominent subjects of the training. Participants circled the chosen answer. Each question had four alternatives, scoring was dichotomous (correct/false). A correct response was coded as 12, and a false response as 1, to be able to combine the VK and MK scales. This might have resulted in the inflation of the effect of VK relative to MK, which should be taken into account when interpreting the results.

Skill

Skill observation is very labor intensive. An alternative, often used method consists of questions about how confident trainees are that they can perform several relevant behaviors. This is referred to as *self-efficacy*: one's perceived performance capability for an activity (Bandura, 1977a). Self-efficacy has been shown to be related both to near-transfer (transfer to a task identical to the trained task), and far transfer (to a task that is an generalization or adaptation of the trained task) (Beier & Kanfer, 2010); and in a variety of domains, such as job interviews (Stumpf, Brief, & Hartman, 1987), sports (Hepler & Chase, 2008), and business and retail (Jawahar, Meurs, & Ferris, 2008). Fourteen self-efficacy questions were developed according to Bandura (2003), covering relevant aspects of the training. Participants indicated per statement how certain they were that they could perform the behavior, by writing down a number between 0 'cannot do at all' and 100 'highly certain can do'.

Affect / attitude

Kraiger and colleagues (1993) divided affect in attitudinal and motivational outcomes. This study measured attitudinal outcomes. Relevant for CRM is an attitude that will make

people use, and elaborate on what they have been taught. Participants indicated how they felt about nine attitude statements on a 12-point Likert scale marked 'strongly disagree', 'slightly disagree', 'neutral', 'slightly agree', to 'strongly agree'. Possible scores ranged from 1 (strongly disagree) to 12 (strongly agree).

Stronger attitudes have a much more profound influence on behavior than weaker ones (Boninger, Krosnick, Berent, & Fabrigar, 1995; Krosnick & Petty, 1995). Attitude strength is a multidimensional construct that predicts transfer and stability of attitudes (Ford et al., 2010; Krosnick & Abelson, 1992; Krosnick & Petty, 1995), of which attitude *extremity* and *importance* were measured. Attitude extremity refers to the degree of favorableness of someone's evaluation, the farther from neutral, the more extreme. The 12-point Likert scale for attitude incorporated the extremity measure, as the distance from the average score. Attitude importance refers to the degree to which a person considers an attitude to be personally important to him or her (Krosnick & Abelson, 1992). Participants indicated this, for each attitude statement, on a separate 12-point Likert scale marked 'not important', 'slightly important', and 'very important'. Possible scores: 1 (not important) to 12 (very important).

Method

Design

The first hypothesis was tested using a within-subjects, pretest-posttest, treatment-only design, see Figure 1. Due to the impossibility to not train participants, no control condition was used. Three main quantitative dependent variables were measured pretest-posttest: *knowledge*, *skill*, and *attitude*. Additionally, *reactions* were measured posttest only.

The second hypothesis was tested using a between-subjects, pretest-posttest, one-way experimental design with random assignment, see Figure 1. One dichotomous independent variable was manipulated, namely BMT (yes/no). Three main quantitative dependent

variables were measured pretest-posttest: *knowledge*, *skill*, and *attitude*. Additionally, *reactions* were measured posttest only.

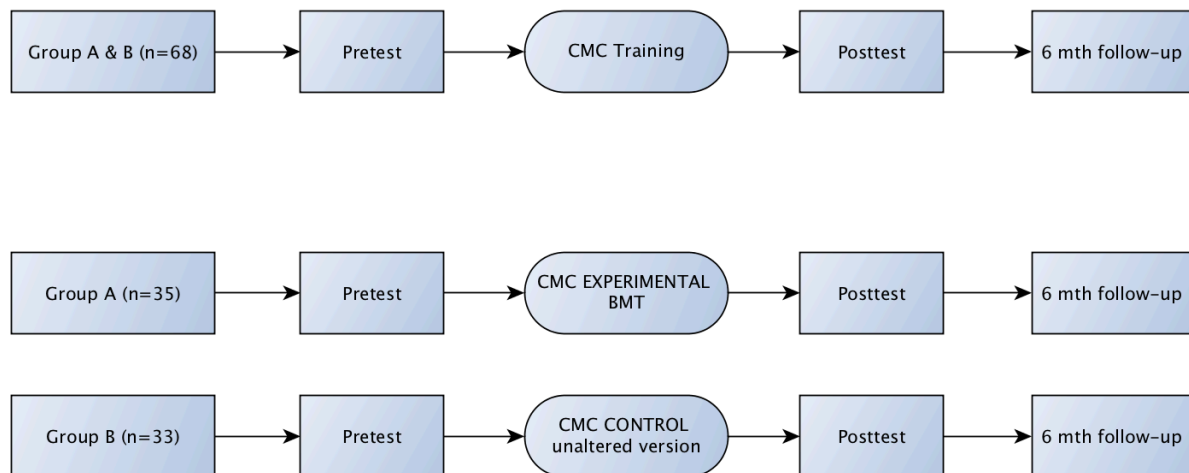


Figure 1: Hypothesis 1 & 2

Participants

68 KLM pilots were scheduled for the CMC training by the scheduling department; all were included in the experiment, 2 females and 66 males. Two CMC-1 and nine CMC-2 courses were investigated, 14 and 54 participants respectively. The 14 CMC-1 participants were inexperienced; the 54 CMC-2 participants had a minimum of 3 years' experience within KLM as first-officers. Six participants were instructors and one participant was also a manager. CMC-1 participants' age was between 22 and 31, average 25.8; not counting two military pilots of 47 and 51 years old. CMC-2 participants' age was between 28 and 47, average 32.8. Selection was semi-random, free from bias: The planning-computer indicated who was due for training. CMC-1 and CMC-2 participants differed on age and experience. The two CMC-1 courses were balanced over experimental conditions. CMC-2 courses were randomized and balanced over experimental conditions, resulting in 4 BMT and 5 non-BMT groups. 35 participants received the experimental BMT training, 33 participants received the control, non-BMT training.

Each course was conducted by two instructors. In the initial design, two dedicated pairs of instructors were planned to conduct all experimental courses. This proved practically impossible due to scheduling and illness. Eventually, the 11 courses were conducted by a total of 12 instructors, randomly assigned. Examination revealed that the distribution of instructors over the BMT and control condition was more or less balanced, and experienced and non-experienced instructors were equally represented in both conditions.

The study was conducted according to the guidelines of Leiden University and approval was obtained from the Leiden University ethics committee. Authorization to conduct the study, adjust the courses, and use the data, was obtained from the Head of Training at KLM. Participants were required to attend the CMC courses as part of the company's training program. No compensation for participating in the study was given.

Materials

To assess learning, the earlier mentioned, pencil-and-paper questionnaire was used.

Procedure

The 3-day CMC training courses were conducted on KLM premises at Schiphol, in a classroom with tables, chairs, and flip-over. A projector was used to support the training material with a PowerPoint presentation. Instructors were briefed by the experimenter before the training. For BMT, instructors were asked to: (a) encourage trainees to set their own goals and make them specific. (b) Model both effective and non-effective behavior. (c) Encourage active practice, increase the time available for it, and encourage new approaches. (d) Provide clear, actionable feedback, related to goals, and (e) explicitly make the connection to the job. They were given a written summary of the main ideas of BMT, including a step-by-step guide, see Appendix A. For the control condition, instructors were similarly briefed, but asked to make sure the BMT aspects were absent. They also received a summary, see Appendix B.

A participant information and consent form informed participants of the experiment, confidentiality, and voluntariness. The pre-questionnaire was filled out on location, before training. To prevent social desirability, participants were seated separated from each other. After training, participants filled out the post-questionnaire, again seated separately. A debriefing form gave information about the purpose of the study, its methodology, and BMT. Instructors were briefed on how to administer the informed consent, debriefing, and questionnaires. Participants will receive a follow-up questionnaire six months after the training. The results of the follow-up study are not included in this report, because the data were not available at the time of writing.

In both conditions, four tasks were practiced during CMC-1 and -2 training. (1) Active listening, a technique called Listen, Ask, Summarize (LAS). (2) Providing feedback on undesired behavior, a technique consisting of: verbalizing undesired behavior, verbalizing the effect on oneself, providing suggestions for desired behavior. (3) Passenger address (PA): giving informative, welcoming, and reassuring PA talks. (4) Handling interpersonal conflicts: Participants practiced five styles for handling conflicts. Before practice, a general, discussion-based introduction into the subject, and after practice a short summary was given. The four experimentally manipulated practice tasks were a subpart of the CMC training. Time devoted to the experimental content was approximately one third (1/3) of the total training time, i.e. \pm 6-8 hours. The order of tasks was the same in all courses.

Analysis

Outliers. The raw data was inspected for outliers on the item, subscale, and scale level. The criterion was a difference of more than 2.5 SDs from the mean (Meyers, Gamst, & Guarino, 2006), but also that this occurred only on one or two items. Low or high scores on most items, but a subscale score within 2.5 SDs was not sufficient for classification as outlier.

Missing values. After the first course some small changes were made to the questionnaire: MK item 4, 5, 7, and 10, and AT item 1 were added. This resulted in missing values for these variables for 4 participants. Due to a printing error, MK item 4 was not included in the pre-questionnaire for the second course, resulting in missing values for 7 more participants; a total of 11 missing. Due to this large percentage of missing values for MK4 (16,2%; or 33% as all missing values fell in the non-BMT condition), imputation was unreasonable and MK item 4 was excluded from analysis (Meyers et al., 2006).

The missing values for MK 5, 7, 10, and AT 1, were categorized as missing at random (MAR); probability of being missing unrelated to observed values on other variables, controlling for course number (Meyers et al., 2006; Rubin, 1976; Sterne et al., 2009). Expectation maximization (EM) imputation was used to replace missing values with maximum likelihood estimations (Meyers et al., 2006). In this case primarily to maintain scale integrity: Because item data was averaged to compute scale scores, missing values would make it unclear in which direction a scale might shift. To avoid inflating results, the estimations were derived from scores of BMT and non-BMT participants combined. Confirmatory analyses were repeated post-hoc on the un-imputed data and confirmed that no bias was introduced. Two participants did not fill out the reactions part. It could not be ruled out that these missing values were non-random, no imputation was used here.

Data calculations. After imputation, pre-, post-, and thentest scores were averaged. This yielded mean subscale pre-, post-, and thentest scores for MK, SE, and AT, and pre- and posttest scores for VK and AI. Pre- and posttest scores for the dependent variables KSA were calculated by taking the average of VK and MK subscale scores, and AT and AI subscale scores. SE subscale scores were used as is.

Statistical method. A repeated-measures multivariate analysis of variance (MANOVA) was used to test the hypotheses, with the three dependent variables KSA, the

within-subjects factor time (pre/post), and between-subjects factor BMT (yes/no). For the first hypothesis the main effect of time was interpreted, and for the second hypothesis the interaction effect of time and BMT. The significant effect of time was further analyzed with paired-sample t-tests. On the reactions data, a one-sample t-test was performed for hypothesis 1, and an independent-samples t-test for hypothesis 2. A criterion α level of 0.05 was used throughout the analyses.

Thentests tend to inflate effects, especially for variables with a social desirability component and specific behaviors targeted by the intervention (Hill & Betz, 2005). Therefore, pre-post comparison was used for confirmatory analyses. Thentest subscale scores were only used in an exploratory way to estimate response-shift bias, and social desirability.

Tested assumptions and criteria. Specific MANOVA assumptions were checked (Meyers et al., 2006): sufficient correlation between dependent variables with Bartlett's test of sphericity; independence of participants was covered by random assignment to conditions; univariate normality was assessed graphically and statistically for each dependent variable; homoscedasticity was assessed statistically; and linearity between pairs of dependent variables was assessed graphically and statistically.

Results

Questionnaire

Attitude scale. Initial investigation revealed abnormalities for AT items 7, 8, and 9 that were not evident in pilot testing. These statements were supposed to be indicative of a negative attitude. However, covariance with the rest of the items was usually not negative, and internal consistency was unexpectedly low: Cronbach's $\alpha_{\text{pre}} = 0.368$, $\alpha_{\text{post}} = 0.611$, and $\alpha_{\text{then}} = 0.660$. Most participants with high scores on the rest of the items scored high on these as well, while a positive attitude would predict scores to be low. Because it was unclear what

AT items 7, 8, and 9 were measuring, they were removed from analyses, along with their accompanying attitude importance items.

Reliability estimates were calculated for the subscales, before imputation of missing values. Recalculation of Cronbach's alpha for the 6-item AT subscale yielded: $\alpha_{\text{pilot}} = 0.662$, $\alpha_{\text{pre}} = 0.633$, $\alpha_{\text{post}} = 0.747$, and $\alpha_{\text{then}} = 0.773$. Although the pre- and pilot- subscales were slightly below the desired value of 0.70 for good consistency, they were still well above insufficient (< 0.50) (Tavakol & Dennick, 2011). For the 5-item VK subscale no stable reliability estimates could be calculated due to the small number of items. For the 10-item MK subscale Cronbach's alphas were: $\alpha_{\text{pilot}} = 0.761$, $\alpha_{\text{pre}} = 0.772$, $\alpha_{\text{post}} = 0.813$, and $\alpha_{\text{then}} = 0.859$. Recalculation of the 9-item MK subscale after removal of item 4 yielded: $\alpha_{\text{pilot}} = 0.727$, $\alpha_{\text{pre}} = 0.784$, $\alpha_{\text{post}} = 0.808$, and $\alpha_{\text{then}} = 0.840$. For the SE subscale: $\alpha_{\text{pilot}} = 0.874$, $\alpha_{\text{pre}} = 0.856$, $\alpha_{\text{post}} = 0.911$, and $\alpha_{\text{then}} = 0.899$. For the AI subscale: $\alpha_{\text{pre}} = 0.684$, and $\alpha_{\text{post}} = 0.718$. Validity of the questionnaire was not formally assessed after the study, but the expert panel was consulted on the removal of the three AT and the MK items. They felt that removal of these items would not significantly alter the measured construct.

Data screening

Outliers. No outliers were excluded from analyses. Several low scores were found at the item level (> 2.5 SD). However, these participants scored low on the (sub)scale overall, suggesting a different internal standard: Their (sub)scale scores were comparable to the rest of the sample (< 2.5 SD). On the main sub- and scale scores (VK, MK, SE, AT, AI, K, S, and A) no outliers were identified.

Missing values. The percentage of missing values due to missing items for MK 5, 7, 10, and AT 1 was 5.9%. Skipped items accounted for six further missing values, resulting in percentages between 1.5 and 2.9 %, except for AT_{post}, AT_{then}, and AI_{post} items 1, where one skipped item combined with the initial 5.9%, resulted in a total of 7.4% missing. Randomness

of missing values was tested per subscale with Little's MCAR test, reaching significance only when combined with course number and BMT (p 's < 0.05), but not when combined with the dependent variables and difference scales only (p 's > 0.05). No significance was reached when combining subscales with only the dependent variables. Missing value cases were dummy coded and differences on the dependent variables, BMT and group number were checked with t-tests and Chi-square tests for the dummy variable. No significant differences on KSA, or R were found (p 's > 0.05), while missingness correlated significantly with BMT and group number (p 's < 0.01). Results imply that the data is not MCAR, but are consistent with MAR; EM imputation was therefore used to replace missing values (Garson, 2015; Meyers et al., 2006). Two participants did not fill out the reactions subscale, one in each condition. No imputation was use and their reactions data (only) were excluded from analysis.

Response-shift. To assess response-shift, mean differences between then- and posttest, and between pre- and posttest were compared. For all subscales, then-post differences were significantly larger than pre-post differences; see Table 2.

Table 2

Response-shift: Subscale Mean Differences for Thenpost and Prepost Tests

Subscale	Means		Response-shift	<i>t</i>	<i>df</i>
	Thenpost difference	Prepost difference			
Metacognitive knowledge	1.42 (0.98)	0.90 (0.78)	0.52 (0.90)	4.72***	67
Self-efficacy	8.45 (4.92)	4.66 (5.08)	3.79 (5.22)	5.98***	67
Attitude	1.15 (0.73)	0.26 (0.91)	0.88 (1.00)	7.33***	67

Note. *** = $p \leq .001$. Standard Deviations in parentheses below means.

Quality assurance. Distributions of subscale scores and dependent variables KSA, and R were found to be approximately normal for the dependent variables KSA, and for most

subscales (Shapiro-Wilk p 's > 0.01). Only the VK scales, ATpost, ATthen, and R statistically violated normality. Examination of histograms and normal probability plots revealed that departure from normality was mild at worst.

To check randomization, pretest data of BMT and non-BMT groups were compared. Independent samples t-tests were conducted on gender, age, flying background, aircraft type, crew position, years of experience, and on the five pre-subscales: VK_{pre}, MK_{pre}, SE_{pre}, AT_{pre}, and AI_{pre}, with BMT (yes/no) as grouping variable. No differences between experimental groups were found (p 's $> .15$).

Tests

Manipulation checks. No formal manipulation check was performed. However, after each course the instructors were contacted and they reported no problems incorporating the BMT elements in the experimental courses, and leaving out the elements of BMT in the control condition.

Confirmatory tests. To test hypothesis 1, that KLM's CMC courses improve participants' KSAs and reactions, and hypothesis 2, that BMT improves KSAs and reactions more than the non-BMT control condition, a repeated-measures MANOVA was performed. A two-level within-subjects factor time (pre/post) and a two-level between-subjects factor BMT (yes/no) were used. All 68 participants were included in the analysis, 35 in the BMT condition, and 33 in the non-BMT condition. Evaluations of the properties of the data determined that necessary statistical assumptions were met. Using Wilks criterion (see Table 3), the main effect of time (in other words, the training) significantly affected the composite dependent variable, confirming hypothesis 1 (Wilks λ , $F [3,64] = 28.61$, $p < .001$, partial $\eta^2 = .57$). The effect was quite pronounced, 57% of the total variance was accounted for by the training effect. Univariate ANOVAs were conducted on each dependent variable to determine the locus of the significant multivariate effect. The main effect of time significantly affected

knowledge, $F [1, 67] = 45.30, p < 0.001$, partial $\eta^2 = .376$; skill, $F [1, 67] = 57.12, p < 0.001$, partial $\eta^2 = .460$; and attitude, $F [1, 67] = 9.67, p < 0.003$, partial $\eta^2 = .126$. Means and standard deviations for the main dependent variables' effects are presented in Table 4. Paired samples t-tests were conducted to examine the effects per subscale, all effects were significant (p 's < 0.05), see Table 5 and Figure 2. Cohen's d 's were calculated to compare subscale effects.

For reactions, 2 participants were excluded from analysis due to missing values, one in each experimental condition. 66 participants were included. Evaluations of the properties of the data (normality and linearity) determined that necessary statistical assumptions were met. For hypothesis 1, a one-sample t-test showed that reactions ($M = 9.89, SD = .90$) were significantly above the 6.5 neutral point, $t (65) = 30.68, p < .001$, confirming hypothesis 1: CMC training leads to positive reactions. The effect is quite pronounced, the mean reaction was 3.39 units above neutral on a 12-point scale, i.e. above the 75th percentile.

Hypothesis 2 was disconfirmed: The interaction effect of time and BMT did not significantly affect the composite dependent variable (Wilks $\lambda, F [3,64] = 0.18, p = .91$, partial $\eta^2 = .01$), see Table 3. As no differential effect of BMT training vs non-BMT training was apparent on the multivariate composite of KSA, specific comparisons were not performed. Means and standard deviations for the subscales are summarized in Table 6 and Figure 3 for reference. Furthermore, an independent samples t-test showed that BMT participants ($M = 10.05, SD = .69$) did not report more positive reactions than non-BMT participants ($M = 9.73, SD = 1.06$), $t (64) = 1.48, p = .143$. Although reactions tended to be slightly more positive in the BMT group, BMT training did not lead to significantly more positive reactions than non-BMT training.

Table 3

Time x BMT Repeated Measures Multivariate Analysis of Variance

Source	<i>df</i>	<i>F</i>	Partial η^2	<i>p</i>
Time	(1, 66)	28.61	.57	< .001***
BMT	(1, 66)	.81	.04	.49
Time x BMT	(1, 66)	.18	.01	.90

Note. *** = $p \leq .001$.

Table 4

Mean Scale Scores and Training Effect for All Participants (hypothesis 1)

Subscale	Time		Difference
	Pre	Post	
Knowledge	7.61 (1.22)	8.76 (1.25)	1.15*** (1.50)
Skill	73.71 (7.29)	78.37 (6.61)	4.66*** (5.08)
Attitude	9.72 (1.13)	10.07 (0.93)	0.36** (0.95)

Note. ** = $p \leq .01$, *** = $p \leq .001$. Standard Deviations in parentheses below means.

Table 5

Subscale Pre & Post Scores and Differences (Hypothesis 1)

Subscale	Training		Difference	<i>t</i>	<i>df</i>	<i>d</i>
	Pre	Post				
Verbal knowledge	6.79 (2.21)	8.20 (2.44)	1.41 (2.85)	4.07***	67	0.61
Metacognitive knowledge	8.42 (0.99)	9.32 (0.83)	0.90 (0.78)	9.55***	67	0.99
Self-efficacy	73.71 (7.29)	78.37 (6.61)	4.66 (5.08)	7.56***	67	0.67
Attitude	9.88 (1.17)	10.14 (1.02)	0.26 (0.91)	2.36*	67	0.24
Attitude importance	9.55 (1.24)	10.01 (0.97)	0.45 (1.15)	3.26**	67	0.41

Note. * = $p \leq .05$, ** = $p \leq .01$, *** = $p \leq .001$. Standard Deviations in parentheses below means.

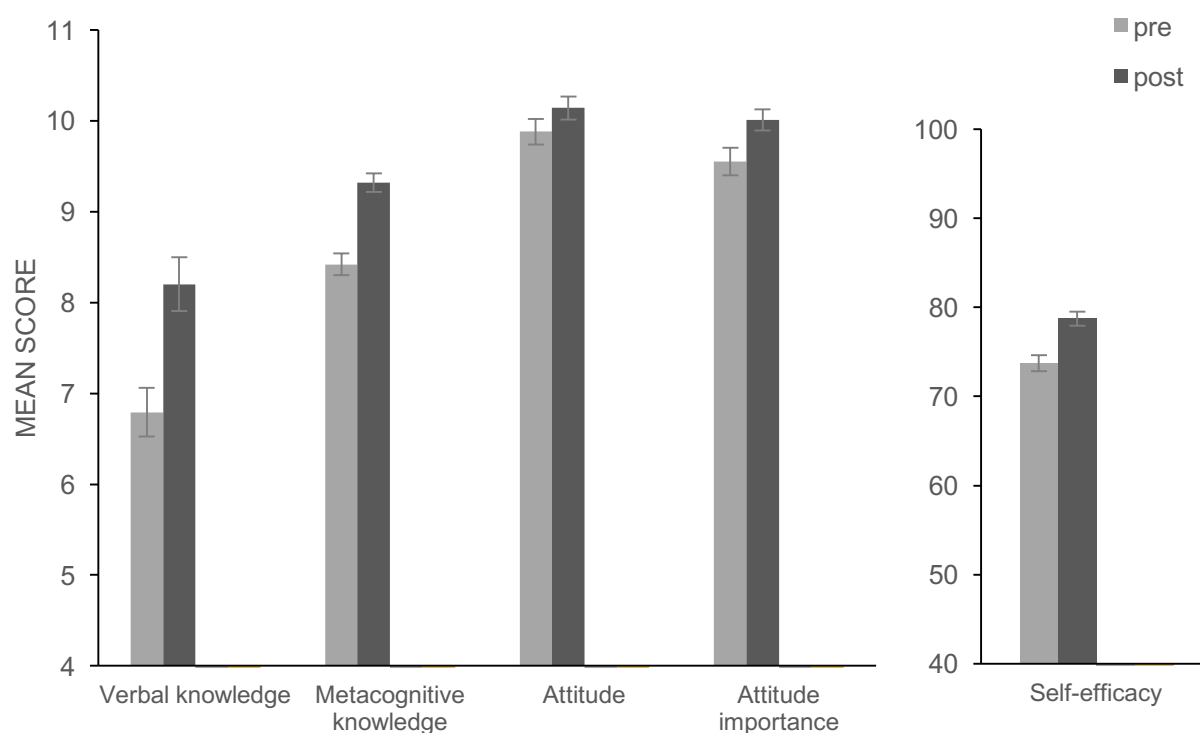


Figure 2. Mean pre- and post-scores per subscale (hypothesis 1). Error bars represent one standard error around the mean.

Table 6

Mean Scale Scores for Experimental Groups (Hypothesis 2)

Scale	Experimental Group						BMT x Time	
	BMT			Non-BMT			Mean diff.	<i>p</i>
	Pre	Post	Diff.	Pre	Post	Diff.		
Knowledge	7.38 (1.19)	8.62 (1.23)	1.24 (1.46)	7.84 (1.21)	8.90 (1.29)	1.06 (1.55)	0.18	n.s.
Skill	73.22 (5.47)	78.24 (5.91)	5.02 (4.36)	74.23 (8.88)	78.24 (7.37)	4.27 (5.80)	0.75	n.s.
Attitude	9.69 (0.99)	10.07 (0.87)	0.38 (0.87)	9.74 (1.27)	10.07 (1.00)	0.33 (1.03)	0.05	n.s.

Note. n.s.= not significant. Standard Deviations in parentheses below means.

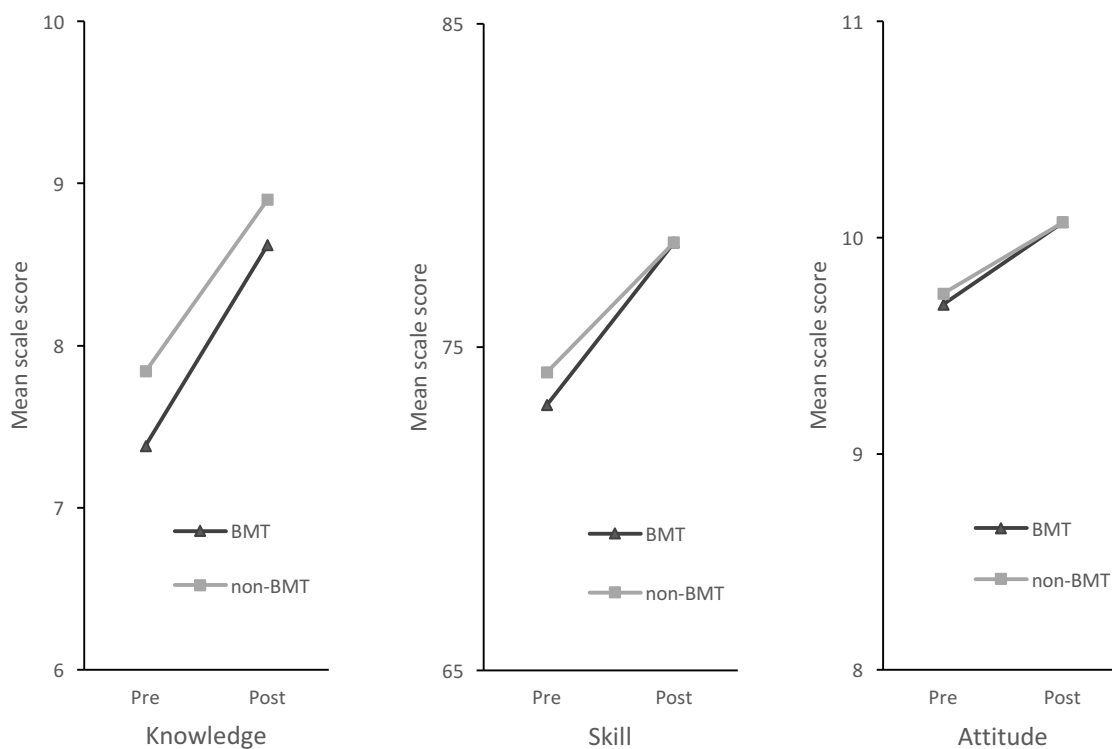


Figure 3. Pre and post mean scale scores per intervention (BMT/non-BMT) (hypothesis 2).

Exploratory tests. To explore if imputation of missing values had biased the results, the main confirmatory tests for hypothesis 1 and 2 were repeated with the original dataset including missing values. These analyses yielded virtually the same results, confirming no

biasing influence of imputation. Hypothesis 1, Wilks λ , $F [3,64] = 28.69$, $p < .001$, partial $\eta^2 = .57$. Hypothesis 2, Wilks λ , $F [3,63] = 0.18$, $p = .91$, partial $\eta^2 = .01$).

The response-shift effect was explored. Using then test scores instead of pretest scores to compute the dependent variables KSA was deemed undesirable, because it would have yielded results composed partly of then post test and partly of prepost test data. Although not computed, because then post differences were larger than prepost differences, see Table 2, it would have resulted in larger learning effects for hypothesis 1. For hypothesis 2, the question was whether response-shift resulted in more then post than prepost improvement for BMT than for non-BMT. Response-shift was calculated by deducting prepost difference from then post difference. T-tests showed that response-shift was not significantly larger in the BMT group than in the non-BMT group (all p 's > 0.50), see Table 7. The improvement in learning in the BMT group compared to the non-BMT group, did not depend on whether it was measured by then post difference, or by prepost difference.

Table 7

Response-shift Differences between Experimental Groups

Subscale	Experimental group		<i>t</i>	<i>df</i>
	BMT response-shift	Non-BMT response-shift		
Metacognitive knowledge	0.55 (0.93)	0.48 (0.88)	0.33 ^a	66
Self-efficacy	4.12 (4.29)	3.44 (6.11)	0.54 ^a	66
Attitude	0.88 (0.84)	0.89 (1.15)	-0.03 ^a	66

Note. ^a = not significant. Standard Deviations in parentheses below means.

Next, differences between BMT and non-BMT groups were explored per subscale. T-tests revealed no significant differences between BMT and non-BMT groups on any of the dependent subscale measures (p 's > 0.05), although metacognitive knowledge approached

significance and tended to be higher in the BMT group ($M = 1.07$, $SD = 0.74$) than in the non-BMT group ($M = 0.72$, $SD = 0.79$), $t(66) = 1.90$, $p = 0.062$, see Table 8. The effect was fairly small however, 0.35 improvement on a 12-point scale.

Table 8

Exploratory Analysis of Subscale Differences between Experimental Groups

Subscale	Experimental group		Difference	<i>t</i>	<i>df</i>
	BMT	Non-BMT			
Verbal knowledge prepost	1.41 (2.64)	1.40 (3.10)	0.01 (1.70) ^b	0.18 ^a	66
Metacognitive knowledge prepost	1.07 (0.74)	0.72 (0.79)	0.35 (0.19) ^b	1.90 ^a	66
Metacognitive knowledge thenpost	1.62 (0.99)	1.20 (0.93)	0.43 (0.23) ^b	1.83 ^a	66
Self-efficacy prepost	5.02 (4.36)	4.27 (5.80)	0.75 (1.24) ^b	0.60 ^a	66
Self-efficacy thenpost	9.14 (3.45)	7.71 (6.08)	1.43 (1.19) ^b	1.20 ^a	66
Attitude prepost	0.25 (0.86)	0.28 (0.98)	-0.03 (0.22) ^b	-0.13 ^a	66
Attitude thenpost	1.13 (0.70)	1.16 (0.78)	-0.04 (0.18) ^b	-0.19 ^a	66
Attitude importance prepost	0.51 (1.09)	0.39 (1.21)	0.13 (0.28) ^b	0.46 ^a	66

Note. ^a = not significant. Standard Deviations in parentheses below means, except ^b: Standard Error of the mean.

Items that BMT would theoretically have the most influence on were explored. Only the first three items of the SE subscale, which had a clear relation to the practice scenario's, were tested in isolation. A t-test showed that they did not show significantly more pre-post improvement for the BMT group ($M = 11.20$, $SD = 7.42$) when compared to the non-BMT group ($M = 8.28$, $SD = 7.41$), $t(66) = 1.62$, $p = 0.110$. The same was done for MK subscale

items with a clear relation to the practice scenario's: 1, 3, 6, 8, 9, and 10. A t-test showed that they improved significantly more from pre to post for the BMT group ($M = 1.08$, $SD = 0.82$) when compared to the non-BMT group ($M = 0.66$, $SD = 0.81$), $t(66) = 2.09$, $p = 0.041$. Not a large effect, but significant nonetheless.

Power

A-priori power estimations were performed for each hypothesis. For hypothesis 1, the smallest estimation of the effect of CRM found in the meta-analysis of O'Connor and colleagues (2008), 0.59, was used. Power was calculated using G*Power (version 3.1) (Faul, Erdfelder, Lang, & Buchner, 2007). With one group, three response variables (KSA), and an alpha level of 0.05, estimated power was 0.98: The recommended a priori estimate of power is $> 70\%$ (Meyers et al., 2006; Stevens, 1980). Post-hoc power calculation on the partial eta squared effect size of .572, yielded a power > 0.99 . For hypothesis 2, the effect of BMT, the power calculation was based on the meta-analytic study of 117 studies by Taylor and colleagues (2005). They found effect sizes of 1.29, 0.91, and 0.28 for declarative knowledge, skill, and attitude respectively. The average of the three was used: 0.827. Power calculation using G*power, 64 participants in two groups of equal size, three response variables, and an alpha level of 0.05, resulted in a power estimate of 0.77. The BMT effect power estimate was recalculated after the study based on the realization that only about 1/3rd of the training time was devoted to content that used BMT. Although the effect of reducing training time is probably non-linear, and effect sizes are not linearly distributed, a conservative estimate of 1/3rd of the initial effect size was used, resulting in a power estimate of 0.13. Post-hoc power calculation yielded an even lower observed power; for the BMT x Time interaction: 0.082.

Discussion

The aim of the study was first, to establish a systematic process of training evaluation at KLM and investigate whether CMC training is achieving its goal in terms of learning and

reactions. And second, to investigate whether introducing BMT training would lead to more positive learning and reactions than standard, unaltered CMC training.

Findings

The first hypothesis was confirmed. KLM's CMC courses significantly improved participants KSAs and reactions. Skill (self-efficacy) and knowledge showed the largest effects, the effect on attitudes was less pronounced. Also, reactions to the training were significantly above neutral. Participants enjoyed it, found it was useful, and indicated that it would influence their behavior and their ability to perform their job. Independent examination of the subscales showed a large effect on metacognitive knowledge, medium-to-large on self-efficacy and verbal knowledge, medium on attitude importance, and small on attitudes. These results do not take correlations between the subscales into account and should therefore be interpreted as relative estimations. The relative sizes of the results were as expected: CMC training targets skill and strategies mostly. They were also fairly consistent with the literature. The present study found a slightly larger effect on knowledge, and smaller on behavior and attitudes than O'Connor and colleagues' (2008) meta-analysis. However, in that study, knowledge concerned verbal knowledge, arguably less relevant for CRM than metacognitive knowledge. This is reflected in our data: The effect of verbal knowledge was less pronounced than that of metacognitive knowledge. The smaller effect on skill that was found could be due to the use of self-report versus behavioral observations. The effect for attitudes was noticeably smaller than that in the meta-analysis. One of the reasons could be that CRM training is well established within KLM and attitudes towards CRM are already quite positive, therefore the training does not influence attitudes as clearly. Also, a shortened questionnaire was used for attitudes, which might have deflated the effect through inadvertent reduction of the breadth of the concept. In summary, taking the specifics of this study into

account, the effects are consistent with the literature, although the effect for attitudes might be slightly smaller.

The second hypothesis was not confirmed. No significant differences were found for participants' KSAs, or reactions between the BMT CMC training and the control condition. Further inspection of the data revealed, that all effects, although not significant, were in the right direction and consistent with the literature (Taylor et al., 2005), except for attitudes. For metacognitive knowledge and self-efficacy, improvement tended to be larger in the BMT condition, just short of significance. For attitudes the differential effect was not significant. Examining the items of the metacognitive and self-efficacy subscales that were especially relevant for the effects of BMT in isolation, revealed that BMT had a significantly more positive effect on the selected metacognitive knowledge items than the control condition. The effect on the selected self-efficacy items was also more positive and approached significance. These examinations were post-hoc and of an exploratory nature. Although no conclusions about the effectiveness of BMT can therefore be derived from these results, they do support the conclusion that the absence of significant differences between the BMT and control condition was not due to the ineffectiveness of BMT, but due to a lack of power. A priori power estimates failed to take into account that the studies in Taylor and colleagues' (2005) meta-analysis, the basis of the estimates, focused completely on BMT; while the present study incorporated BMT only in the practical exercises, about 1/3rd of the training time. The BMT effect was thus severely diluted in this study, which probably resulted in very low power, in spite of the fact that pre-post comparison was used.

Instructors were informally interviewed after each course and reported that BMT training gave them much needed direction. They also reported that participants seemed to like the structure that BMT provided, and its practical way of training. This further supports the assertion that the lack of significant effects is due to a lack of power, as both instructors and

participants appear to find BMT useful, which has been shown to predict actual utility (Alliger, Tannenbaum, & Bennett, 1997).

Limitations

The main limitation of this study is the lack of a control group for hypothesis 1. Without it, no definitive conclusions about the amount of learning from the CMC training can be drawn. However, the effects and their sizes were consistent with the literature's estimations. Furthermore, there was a tendency towards differential effects between the BMT and non-BMT conditions, and they were in the right direction. Combined, this provides some basis for the conclusion that the improvements found were due to training. The sizes of the effects should nevertheless be interpreted with caution, especially taking into account possible differential placebo effects for the dependent measures; attitudes might be more affected than self-efficacy.

The second main limitation of this study is that it relied solely on self-report. Especially the use of self-efficacy to infer skill learning was a shortcoming: although they are correlated, they are not the same. Subjective assessment will be affected by whether trainees liked the training and how well they felt they performed (O'Connor et al., 2002; Stumpf et al., 1987). The use of self-efficacy might thus have overestimated the skill effect. Later efforts should focus on validating the questionnaire with observations, or using observations to assess skill learning.

Another limitation concerns response-shift. It should measure a shift in people's internal scale, but it has been shown that other biases may influence it as well, such as social desirability (Hill & Betz, 2005). When people expect to change and feel that they should have changed on certain dimensions, they are likely to magnify the degree of change. From our data there is no way to tell whether the difference between prepost and thenpost scores is due to the shifting of an internal scale, or due to social desirability. It is likely however, that at

least some of the observed response-shift is due to social desirability. Of course this does not offset the problem of a lack of control group, because there is no way to know if posttest scores were not simply inflated. But asking participants for their test scores might have given them a way to inflate their learning effect without having to resort to inflating posttest scores, thereby decreasing social desirability effects on posttest scores, hopefully making the estimates reported in this study more accurate.

An additional limitation is that although all courses were initially planned to be conducted by the same two instructor pairs, due to practical difficulties this proved impossible. Instructors and participants report that differences between instructors account for much of the observed differences between courses. It might thus well have been that this introduced a significant amount of variability, making the detection of a difference between the BMT and control condition more difficult.

Lastly, regarding the questionnaire that was used, the experimental data showed some odd effects. Being negatively phrased, attitude items 7, 8, and 9 should have yielded low scores for participants scoring high on the rest of the attitude items. However, this was not the case. It appears as if participants either did not realize that the statements were to be answered with disagreement for positive attitude, or that the content of the statements was not perceived as being negative for CRM. Supporting the last explanation is the fact that the items concerned decision making under duress, leaving personal problems at home, and the influence of fatigue. Because the training addressed these subjects, participants might have assumed that they should be able to handle them, and were reluctant to indicate that they were not. As it was unclear what the items were measuring, the items were removed to prevent clouding the interpretability of the results. Even though the expert panel deemed the scope of the attitude items sufficient, the questionnaire might have measured the attitude domain only partly. This is a problem of the instrument more generally: Validity and reliability were

assessed but not as rigorously as is the case with other, more widely used questionnaires.

Therefore, research effort should be directed at developing easy-to-use, reliable, and validated multidimensional instruments.

Conclusion

This study had two main goals. First, to begin establishing a CRM training evaluation program at KLM, and second, to assess the effectiveness of introducing BMT training. Two questions were investigated: (1) Does KLM's CMC training improve participants KSAs and reactions, and (2) does BMT training improve participants KSAs and reactions more than non-BMT training? A questionnaire was developed to measure CRM learning and reactions. Overall, CMC training significantly improved participants KSAs and reactions. However, BMT training was not significantly more effective than non-BMT training. This appeared to be primarily due to a lack of power. Although not significant, BMT training tended to improve participants KSAs more and lead to more positive reactions than non-BMT training. Furthermore, metacognitive knowledge and self-efficacy were theorized to be most strongly affected by BMT training, and results showed that they were, albeit non-significantly so. These results suggest that BMT is a promising training method that could be used to improve CRM training. Also, instructors report that participants seem to like it, and instructors themselves like the structure and guidance that it provides.

This study adds to the body of research that supports the effectiveness of CRM training. However, multilevel evaluation of CRM training and development of effective, validated training programs should continue. The industry is in dire need of such information. Furthermore, effort should be directed at making those programs more widely available. Cooperation between researchers and industry parties could help achieve this goal, making the skies a safer place.

References

- Alliger, G. M., Tannenbaum, S. I., & Bennett, W. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50*(2), 341–358.
- Alvarez, K., Salas, E., & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. *Human Resource Development Review, 3*(4), 385–416.
- Arthur, W., Bennett, W., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology, 88*(2), 234–245.
- Bandura, A. (1977a). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.
- Bandura, A. (1977b). *Social learning theory*. London: Prentice-Hall.
- Bandura, A. (2003). Guide for constructing self-efficacy scales. In *Self-efficacy beliefs of adolescents* (5th ed., pp. 307–337). Greenwich, CT.
- Beier, M. E., & Kanfer, R. (2010). Motivation in training and development: A phase perspective. In S. W. J. Kozlowski & E. Salas (Eds.), *Learning, training, and development in organizations* (pp. 65–97). New York: Taylor & Francis.
- Boninger, D. S., Krosnick, J. A., Berent, M. K., & Fabrigar, L. R. (1995). The causes and consequences of attitude importance. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength* (pp. 159–189). Mahwah, NJ: Lawrence Erlbaum Associates.
- Burke, M. J., & Day, R. R. (1986). A cumulative study of the effectiveness of managerial training. *Journal of Applied Psychology, 71*(2), 232–245.
- Cartwright, T. J., & Atwood, J. (2014). Elementary pre-service teachers' response-shift bias: Self-efficacy and attitudes toward science. *International Journal of Science Education, 36*(14), 2421–2437.
- Civil Aviation Authority. (2002). *Flight crew training: Cockpit resource management (CRM)*

- and line-oriented flight training (LOFT) (CAP 720)*. Cheltenham, Glos.: Documedia Solutions Ltd.
- Civil Aviation Authority. (2003). *Methods used to evaluate the effectiveness of flightcrew CRM training in the UK aviation industry (CAA PAPER 2002/5)*. Cheltenham, Glos.: Documedia Solutions Ltd.
- Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., et al. (2013). Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis. *Medical Teacher*, 35(1), 867–898.
- De Galan, M. (2016). *Cognitive enhancement analysis: Crew resource management training* (Unpublished report). Leiden University, Leiden, The Netherlands.
- De Galan, M. (2017). *CRM Questionnaire development*. (Unpublished report). Leiden University, Leiden, The Netherlands.
- EASA. (2014, April 24). *Acceptable Means of Compliance (AMC) and Guidance Material (GM) to Part-ORO - Issue 2*. Retrieved from: <https://www.easa.europa.eu/document-library/acceptable-means-of-compliance-and-guidance-materials/part-oro-amc-gm-issue-2>
- FAA. (2016, June 30). *Electronic Code of Federal Regulations, §121- Subpart N*. Retrieved from http://www.ecfr.gov/cgi-bin/retrieveECFR?gp=1&SID=8b3b2f7e2cf1179a111d70604082e34c&ty=HTML&h=L&mc=true&n=sp14.3.121.n&r=SUBPART#se14.3.121_1403
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.
- Ford, J. K., Kraiger, K., & Merritt, S. M. (2010). An updated review of the multidimensionality of training outcomes: New directions for training evaluation

- research. In S. W. J. Kozlowski & E. Salas (Eds.), *Learning, training, and development in organizations* (pp. 135–165). New York, NY: Taylor & Francis.
- Garson, G. D. (2015). *Missing values analysis & data imputation*. Asheboro, NC: Statistical Associates Publishing. Available from <http://www.statisticalassociates.com/missingvaluesanalysis.htm>
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112.
- Helmreich, R. L., & Clayton Foushee, H. (2010). Why CRM? Empirical and theoretical bases of human factors training. In B. G. Kanki, R. L. Helmreich, & J. M. Anca (Eds.), *Crew Resource Management* (2nd ed., pp. 3–57). Amsterdam: Elsevier.
- Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The evolution of crew resource management training in commercial aviation. *The International Journal of Aviation Psychology*, 9(1), 19–32.
- Hepler, T. J., & Chase, M. A. (2008). Relationship between decision-making self-efficacy, task self-efficacy, and the performance of a sport skill. *Journal of Sports Sciences*, 26(6), 603–610.
- Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, 26(4), 501–517.
- Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. W., & Gerber, S. K. (1979). Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychological Measurement*, 3(1), 1–23.
- Jawahar, I. M., Meurs, J. A., & Ferris, G. R. (2008). Self-efficacy and political skill as comparative predictors of task and contextual performance: A two-study constructive replication. *Human Performance*, 21(2), 138–157.
- Kirkpatrick, D. I. (1976). Evaluation of training. In R. L. Craig & L. R. Bittel (Eds.), *Training*

- and development handbook* (pp. 18.1–18.27). New York, NY: McGraw Hill.
- Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*(2), 311–328.
- Krosnick, J. A., & Abelson, R. P. (1992). The case for measuring attitude strength in surveys. In J. M. Tanur (Ed.), *Questions about questions* (pp. 177–203). New York, NY: Russell Sage Foundation.
- Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength* (pp. 1–24). Mahwah, NJ: Lawrence Erlbaum Associates.
- Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate data research*. London: Sage.
- Morin, L. (1998). *Mental practice and goal setting as transfer of training strategies: Their influence on self-efficacy and task performance of team leaders in an organizational setting*. (Doctoral dissertation). Retrieved from National library of Canada: <https://tspace.library.utoronto.ca/bitstream/1807/12292/1/NQ35259.pdf>
- Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2015). *Straight Choices* (2nd ed.). New York, NY: Psychology Press.
- O'Connor, P., Campbell, J., Newon, J., Melton, J., Salas, E., & Wilson, K. A. (2008). Crew resource management training effectiveness: A meta-analysis and some critical needs. *International Journal of Aviation Psychology, 18*(4), 353–368.
- O'Connor, P., Hörmann, H. J., Flin, R., & Lodge, M. (2002). Developing a method for evaluating Crew Resource Management skills: A European perspective. *International Journal of Aviation Psychology, 12*(3), 263–285.
- Passmore, J., & Velez, M. J. (2014). Training evaluation. In K. Kraiger, J. Passmore, N.

- Rebelo dos Santos, & S. Malvezzi (Eds.), *The Wiley Blackwell handbook of the psychology of training, development, and performance improvement* (pp. 136–153). Malden, MA: Wiley Blackwell.
- Rosen, M. A., Schiebel, N., Salas, E., Wu, T. S., Silvestri, S., & King, H. B. (2012). How can team performance be measured, assessed, and diagnosed? In E. Salas & K. Frush (Eds.), *Improving patient safety through teamwork and team training* (pp. 59–82). New York, NY: Oxford University Press.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.
- Salas, E., & Rosen, M. A. (2010). Experts at work: Principles for developing expertise in organizations. In S. W. J. Kozlowski & E. Salas (Eds.), *Learning, training, and development in organizations* (pp. 99–134). New York: Taylor & Francis.
- Salas, E., Burke, C. S., & Bowers, C. A. (2001). Team training in the skies: Does crew resource management (CRM) training work? *Human Factors*, *43*(4), 641–674.
- Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Science in the Public Interest*, *13*(2), 74–101.
- Salas, E., Wilson, K. A., Burke, C. S., & Wightman, D. C. (2006). Does crew resource management training work? An update, an extension, and some critical needs. *Human Factors*, *48*(2), 392–412.
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, *19*(4), 460–475.
- Schwartz, C. E., & Sprangers, M. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science & Medicine*, *48*(11), 1531–1548.
- Schwartz, C. E., Sprangers, M. A. G., Carey, A., & Reed, G. (2004). Exploring response shift

- in longitudinal data. *Psychology & Health*, 19(1), 51–69.
- Shuffler, M. L., Salas, E., & Xavier, L. F. (2010). The design, delivery and evaluation of crew resource management training. In B. G. Kanki, R. L. Helmreich, & J. M. Anca (Eds.), *Crew Resource Management* (2nd ed., pp. 205–232). Amsterdam: Elsevier.
- Smith-Jentsch, K. A., Salas, E., & Baker, D. P. (1996). Training team performance-related assertiveness. *Personnel Psychology*, 49(4), 909–936.
- Sprangers, M., & Hoogstraten, J. (1989). Pretesting effects in retrospective pretest-posttest designs. *Journal of Applied Psychology*, 74(2), 265–272.
- Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., et al. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *British Medical Journal*, 338, b2393–b2393. Retrieved from: <http://www.bmj.com/cgi/doi/10.1136/bmj.b2393>
- Stevens, J. P. (1980). Power of the multivariate analysis of variance tests. *Psychological Bulletin*, 88(3), 728–737.
- Stumpf, S. A., Brief, A. P., & Hartman, K. (1987). Self-efficacy expectations and coping with career-related events. *Journal of Vocational Behavior*, 31(1), 91–108.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55.
- Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. L. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology*, 90(4), 692–709.

Appendix A

BMT Summary for Instructors, Including BMT Exercises

Instructie CMC experiment voor instructeurs (BMT)

Bedankt dat jullie mee willen werken aan het experiment om de CMC cursussen te beoordelen en verbeteren. Om het allemaal soepel te laten verlopen hier een stapsgewijze instructie.

1. Eerst de introductie van de dagen.
2. De kandidaten krijgen een korte uitleg over het experiment en moeten dit formulier ondertekenen. Het betreft het 'participant information and consent form'. Mochten er vragen zijn, probeer deze dan zo goed mogelijk te beantwoorden.
3. Zet de kandidaten uit elkaar om te voorkomen dat ze zich bekeken voelen, dat voorkomt sociaal wenselijk antwoorden.
4. De kandidaten vullen de 'CRM questionnaire (pre)' in. Dit kost ongeveer 10-15 minuten. Vergeet niet om per kandidaat zijn of haar volgnummer op het voorblad van de questionnaire te schrijven, zonder dit nummer kunnen de data niet verwerkt worden!
5. Probeer eventuele inhoudelijke vragen over de verbale kennis-vragen in de questionnaire nog even niet te beantwoorden. Deze vragen zullen na afloop van de cursus nog een keer gesteld worden (post). Mochten ze tijdens de cursus behandeld worden en zo worden beantwoord, dan is dat natuurlijk geen punt.
6. Na afloop van de cursus (eind van dag 3), vullen de kandidaten de 'CRM questionnaire (post)' in. Zet ook hiervoor de kandidaten uit elkaar. Deze vragenlijst stelt de kandidaten voor een aantal items twee vragen. Een vraag over hoe ze zichzelf nu inschatten en een vraag over hoe ze zichzelf, als ze nu terug kijken, voor de cursus zouden hebben beoordeeld. De bedoeling daarvan is dat we een idee krijgen of hun interne standaard is veranderd: oftewel of ze de lat na de cursus misschien hoger leggen.
7. Vergeet ook hier niet de volgnummers per kandidaat op de questionnaire te schrijven!
8. Geef *na* het invullen van de vragenlijst de 'debriefing' formulieren aan de kandidaten.

De ingevulde vragenlijsten kun je in een envelop stoppen en op kantoor achterlaten, laat wel even weten waar je ze neerlegt.

BMT variant

De cursus die jullie de komende dagen geven is de 'experimentele', Behavior Modeling Training (BMT) variant. Een korte samenvatting om ervoor te zorgen dat de componenten van BMT de komende dagen ook echt aanwezig zijn:

1. Voor het oefenen worden de doelen van de oefening specifiek kenbaar gemaakt (bij LSD bijvoorbeeld: luister, vat samen, vraag door; zie instructie per onderwerp)
2. De oefening wordt eerst voorgedaan door de instructeurs. Het voordoen is inclusief korte feedback. Dit demonstreert hoe de kandidaten het oefenen daarna aan moeten pakken, dat maakt het makkelijker.
3. Er wordt zoveel mogelijk geoefend: dwz doe de oefeningen met alle kandidaten. Probeer ze zo kort mogelijk te houden, het is beter nog een keer te oefenen, dan helemaal uit te diepen waarom iets wel of niet goed ging. Leren door te ervaren!
4. Na afloop van de oefeningen wordt feedback gegeven. Feedback is gestructureerd aan de hand van de concrete doelen die in stap 1 kenbaar zijn gemaakt.

Ik begrijp dat het soms lastig zal zijn om alle elementen te includeren, te demonstreren en veel te oefenen, maar houdt in gedachten dat het de enige manier is om erachter te komen of het echt werkt.

Tenslotte wens ik jullie heel veel plezier met het geven van de cursus! Mocht je nog vragen hebben, aarzel dan niet om even te bellen: 06-55818646.

Met vriendelijke groeten, Michiel de Galan

BMT instructies LSD

1. Introductie onderwerp
2. Doelen expliciet maken voor het oefenen (zet ze evt op een flip-over)
 - Is er **geluisterd**?
 - Is er **samen gevat**?
 - Is er **doorgevraagd**?
 - Is dat op een empathische **manier** gedaan? (Heeft ondervraagde het gevoel dat er geluisterd is?)
 - Is **inhoud** duidelijk geworden?
3. Modelleren
 - Instructeurs demonstreren een LSD oefening. Houd het kort.
 - Instructeurs geven elkaar feedback. Aan de hand van doelen, geef concrete feedback over hoe anders had kunnen doen om doelen te behalen. Niet te lang ingaan op waarom iets niet werkte.
4. Oefenen
 - Verdeel kandidaten in 2 groepen.
 - Kandidaten gaan 1 op 1 oefenen, eerst 1 paar, dan andere paar. Met vooraf vastgesteld onderwerp (zelf of door instructeur). Korte sessies, begeleiding door instructeur.
 - Geef elkaar feedback, onder begeleiding van instructeur. Aan de hand van doelen, geef concrete feedback over hoe anders had kunnen doen om doelen te behalen. Niet blijven hangen in waarom iets niet werk en waarom dat gedaan is.
 - Nogmaals oefenen, als oefening voldoende duidelijk is, kan dat gelijktijdig in groepjes van 2 (1 op 1), terwijl instructeur rondloop en begeleidt.
 - Geef korte feedback, aan de hand van doelen, concreet op wat te veranderen om doelen te behalen.
5. Nabespreken en link naar werksituatie maken
 - Nabespreken, aan de hand van leerdoelen van kandidaten.
 - Maak link naar werksituatie: hoe ga je dat nu op de lijn gebruiken?

BMT instructies HEG

1. Introductie onderwerp
2. Doelen expliciet maken voor het oefenen (zet ze evt op een flip-over)
 - **Diagnosticeren** welke **handeling** het probleem geeft bij mij.
 - **Vertalen** wat daarvan het **effect** op mij is.
 - Dat op een niet aanvallende (objectieve) manier **aangeven** aan feedback-ontvanger (het criterium is hier de ervaring van de ontvanger).
 - Eventueel aangeven van **gewenst** gedrag (niet aanvallend).
3. Modelleren
 - Instructeurs demonstreren een HEG oefening. Houd het kort.
 - Instructeurs geven elkaar feedback. Aan de hand van doelen, geef concrete feedback over hoe anders had kunnen doen om doelen te behalen. Niet te lang ingaan op waarom iets niet werkte.
4. Oefenen
 - Verdeel kandidaten in 2 groepen.
 - Kandidaten gaan 1 op 1 oefenen, eerst 1 paar, dan andere paar. Met vooraf vastgesteld onderwerp (zelf of door instructeur). Korte sessies, begeleiding door instructeur.
 - Geef elkaar feedback, onder begeleiding van instructeur. Aan de hand van doelen, geef concrete feedback over hoe anders had kunnen doen om doelen te behalen. Niet blijven hangen in waarom iets niet werk en waarom dat gedaan is.
 - Nogmaals oefenen, als oefening voldoende duidelijk is, kan dat gelijktijdig in groepjes van 2 (1 op 1), terwijl instructeur rondloop en begeleidt.
 - Geef korte feedback, aan de hand van doelen, concreet op wat te veranderen om doelen te behalen.
5. Nabespreken en link naar werksituatie maken
 - Nabespreken, aan de hand van leerdoelen van kandidaten.
 - Maak link naar werksituatie: hoe ga je dat nu op de lijn gebruiken?

BMT instructies Conflict Hantering

1. Introductie onderwerp
2. Doelen expliciet maken voor het oefenen (zet ze evt op een flip-over)
 - **Proberen** van verschillende conflict-**stijlen**.
 - Inzicht in het feit dat verschillende stijlen op een zelfde case **toepasbaar** zijn.
 - Inzicht in dat elke stijl verschillende **gevolgen** heeft.
 - Realiseren dat je een stijl kunt **kiezen** op basis van wat je op dat moment het beste acht (flexibiliteit).
3. Modelleren
 - Instructeurs demonstreren een conflict oefening. Houd het kort.
 - Instructeurs geven elkaar feedback. Aan de hand van doelen, geef concrete feedback over hoe anders had kunnen doen om doelen te behalen. Niet te lang ingaan op waarom iets niet werkte.
4. Oefenen
 - Verdeel kandidaten in 2 groepen.
 - Gebruik duidelijke cases die door de kandidaten zelf zijn verzonnen. Ter ondersteuning zijn tevens pre-made cases beschikbaar die als leidraad/oefening gebruikt kunnen worden (zie map op Google Drive: MyDrive/Human Factors Group/CMC-2/Oefeningen).
 - Kandidaten gaan 1 op 1 oefenen, eerst 1 paar, dan andere paar. Korte sessies, begeleiding door instructeur.
 - Geef elkaar feedback, onder begeleiding van instructeur. Aan de hand van doelen, geef concrete feedback over hoe anders had kunnen doen om doelen te behalen. Niet blijven hangen in waarom iets niet werk en waarom dat gedaan is.
 - Nogmaals oefenen, als oefening voldoende duidelijk is, kan dat gelijktijdig in groepjes van 2 (1 op 1), terwijl instructeur rondloop en begeleidt.
 - Geef korte feedback, aan de hand van doelen, concreet op wat te veranderen om doelen te behalen.
5. Nabespreken en link naar werksituatie maken
 - Nabespreken, aan de hand van leerdoelen van kandidaten.
 - Maak link naar werksituatie: hoe ga je dat nu op de lijn gebruiken?

BMT instructies PA

1. Introductie onderwerp

2. Doelen expliciet maken voor het oefenen (werk ze uit met groep en zet op een flip-over)

A) Vanuit cockpit

De (inhoudelijke) doelen halen we uit de ontvangen informatie van de Communicatie Afdeling van KLM:

- Doelen PA: 3 clusters
 1. Informatievoorziening
 2. Gastheerschap
 3. Geruststellen
- Werk dit uit met groep en schrijf ze op flip-over.

Besteed verder aandacht aan:

- Logische opbouw PA.
- Vorbereiding PA.
- Oefenen PA (evt, bv subvoce)
- Duidelijk spreken (verstaanbaar, rustig, niet te snel).
- Overbrengen boodschap. Let op 3 clusters.
- Omgaan met eventuele tegenslagen.
- Afronding, geven van vooruitzicht.

B) Voor groep (in cabine)

Veel is hetzelfde als bij A, maar hier gaat het vooral om het ervaren van wat spreken voor een groep met je doet:

- Omgaan met spanning, angst en gepieker (vooraf & tijdens).
- Contact maken met publiek.
- Duidelijk spreken (verstaanbaar, rustig, niet te snel).
- Overbrengen boodschap en afronding.
- Omgaan met eventuele tegenslagen.
- Gelegenheid geven voor reacties/vragen van publiek (evt.).

3. Modelleren

- Instructeurs demonstreren een PA. Houd het kort.
- Instructeurs geven elkaar feedback. Aan de hand van doelen, geef concrete feedback over hoe anders had kunnen doen om doelen te behalen. Niet te lang ingaan op waarom iets niet werkte.

4. Oefenen

A) vanuit cockpit:

- Verdeel kandidaten in 2 groepen.
- Kandidaten gaan per groep oefenen, bereidt PA voor en geef 'm aan tafel of neem op op telefoon. Met vooraf vastgesteld onderwerp (zelf of door instructeur). Korte sessies, begeleiding door instructeur.
- Geef elkaar feedback, onder begeleiding van instructeur. Aan de hand van doelen, geef concrete feedback over hoe anders had kunnen doen om doelen te behalen. Niet blijven hangen in waarom iets niet werk en waarom dat gedaan is.
- Iedereen doet minimaal 1 PA in groepje. Liefst een 2e keer, dit kan als oefening kort en bondig gehouden wordt.

ZOZ

- B) voor groep
 - Nogmaals PA doen. Nu voor groep, met focus op de doelen onder B.
 - Geef korte feedback, aan de hand van doelen, concreet op wat te veranderen om doelen te behalen.
- 5. Nabespreken en link naar werksituatie maken
 - Nabespreken, aan de hand van leerdoelen van kandidaten.
 - Maak link naar werksituatie: hoe ga je dat nu op de lijn gebruiken?

Appendix B

Non-BMT Summary for Instructors

Instructie CMC experiment voor instructeurs (non-BMT)

Bedankt dat jullie mee willen werken aan het experiment om de CMC cursussen te beoordelen en verbeteren. Om het allemaal soepel te laten verlopen hier een stapsgewijze instructie.

1. Eerst de introductie van de dagen.
2. De kandidaten krijgen een korte uitleg over het experiment en moeten dit formulier ondertekenen. Het betreft het 'participant information and consent form'. Mochten er vragen zijn, probeer deze dan zo goed mogelijk te beantwoorden.
3. Zet de kandidaten uit elkaar om te voorkomen dat ze zich bekeken voelen, dat voorkomt sociaal wenselijk antwoorden.
4. De kandidaten vullen de 'CRM questionnaire (pre)' in. Dit kost ongeveer 10-15 minuten. Vergeet niet om per kandidaat zijn of haar volgnummer op het voorblad van de questionnaire te schrijven, zonder dit nummer kunnen de data niet verwerkt worden!
5. Probeer eventuele inhoudelijke vragen over de verbale kennis-vragen in de questionnaire nog even niet te beantwoorden. Deze vragen zullen na afloop van de cursus nog een keer gesteld worden (post). Mochten ze tijdens de cursus behandeld worden en zo worden beantwoord, dan is dat natuurlijk geen punt.
6. Na afloop van de cursus (eind van dag 3), vullen de kandidaten de 'CRM questionnaire (post)' in. Zet ook hiervoor de kandidaten uit elkaar. Deze vragenlijst stelt de kandidaten voor een aantal items twee vragen. Een vraag over hoe ze zichzelf nu inschatten en een vraag over hoe ze zichzelf, als ze nu terug kijken, voor de cursus zouden hebben beoordeeld. De bedoeling daarvan is dat we een idee krijgen of hun interne standaard is veranderd: oftewel of ze de lat na de cursus misschien hoger leggen.
7. Vergeet ook hier niet de volgnummers per kandidaat op de questionnaire te schrijven!
8. Geef *na* het invullen van de vragenlijst de 'debriefing' formulieren aan de kandidaten.

De ingevulde vragenlijsten kun je in een envelop stoppen en op kantoor achterlaten, laat wel even weten waar je ze neerlegt.

Non-BMT variant

De cursus die jullie de komende dagen geven is de 'standaard', non Behavior Modeling Training (BMT) variant. Een korte samenvatting om ervoor te zorgen dat de componenten van BMT de komende dagen ook echt afwezig zijn:

1. Voor het oefenen worden de doelen van de oefening NIET specifiek kenbaar gemaakt (uiteraard wel gewoon de introductie van het onderwerp)
2. De oefening wordt NIET voorgedaan door de instructeur. De kandidaten gaan direct zelf aan de slag.
3. Er wordt NIET meer geoefend dan normaal: dwz de oefeningen die aan bod komen worden gedaan, maar niet meerdere keren herhaald en de cases voor conflict hantering worden NIET gebruikt.
4. Na afloop van de oefeningen wordt op de normale manier feedback gegeven. Feedback is NIET gestructureerd (want er zijn geen concrete doelen kenbaar gemaakt, zie stap 1).

Ik begrijp dat het soms lastig zal zijn om sommige elementen achterwege te laten, maar houdt in gedachten dat er ook andere cursussen langs gaan komen die juist de nadruk daarop gaan leggen. Om erachter te komen of dat echt werkt is dit de enige manier.

Tenslotte wens ik jullie heel veel plezier met het geven van de cursus! Mocht je nog vragen hebben, aarzel dan niet om even te bellen: 06-55818646.

Met vriendelijke groeten, Michiel de Galan