Air Force Synthetic Training Effectiveness Research in the Australian Context

Christopher Best, Dawei Jia, and Graeme Simpkin
Aerospace Division, Defence Science and Technology Organisation
Fishermans Bend, Victoria
AUSTRALIA

christopher.best@dsto.defence.gov.au / dawei.jia@dsto.defence.gov.au /
graeme.simpkin@dsto.defence.gov.au

ABSTRACT

A high degree of individual skill is a necessary, but not sufficient condition for success in air combat. It is for this reason that the performance of teams and teams-of-teams is of great concern for all air forces. Live training exercises can provide excellent opportunities to learn the knowledge, skills, and attitudes required to act as part of a coordinated fighting force. However, live exercises are expensive and logistically challenging. The tools and methods of synthetic collective training (SCT; i.e., networked simulators and associated support systems) provide a potential means by which to address some of the shortcomings of live training. The objective of the Black Skies series of exercises (EBS) is to provide an empirical basis for understanding how best to design, manage, and support SCT for air combat. The outcomes presented here suggest that the systems and methods used during EBS constitute a solid baseline for development. However, several areas requiring further research are also identified.

1.0   INTRODUCTION

A high degree of individual skill is a necessary, but not sufficient condition for success in air combat. Success in this domain also depends on effective coordination between operators across a range of specialised roles. It is for this reason that the performance of teams and teams-of-teams is of great concern for all air forces. Significant resources are invested by air forces in training aimed at improving the ability of operators to act effectively as members of a coordinated force in operationally realistic mission scenarios. Live training involving large numbers of personnel and airborne assets can provide excellent learning opportunities. However, live exercises are expensive and logistically challenging. Environmental, regulatory (e.g., airspace), and safety constraints also place bounds on the kinds of learning experiences that can be provided during live training.

The tools of synthetic training (i.e., simulators and associated support systems) provide a potential means by which to address some of the shortcomings of live training. Recognition of this potential has led to significant investments in capability by Air Forces in many nations. To ensure the greatest possible return on investment, the development of such capability has typically been guided by the outcomes of integrated programs of research and development. For example, investigations conducted by the Air Force Research Laboratory (AFRL) [1, 2] and the Defence Science and Technology Laboratory (Dstl) [3, 4] have been instrumental in understanding the technical infrastructure required to simulate realistic air-combat missions and the methods by which synthetic training events should designed and managed. While research and development in this area has been underway for almost two decades, it is only more recently that the potential of this kind of training has been recognised in Australia.

1.1   The Australian Context: Exercise Black Skies (EBS)

Under the direction of Royal Australian Air Force (RAAF) Headquarters, in 2004 the Defence Science and Technology Organisation (DSTO) established a program of collaborative research with AFRL aimed at better understanding SCT effectiveness. As part of this collaboration, DSTO and AFRL conducted the Pacific Link exercises [5, 6]. These events represented the first time that air force operators in Australia and the United States had participated together in realistic, air-combat missions via a distributed simulation network. While this was a valuable first step, a shortcoming of these activities was that they were conducted using unclassified systems, networks, and scenarios. This led to some doubts about the generalisability of the findings arising from them. To remediate this, DSTO conducted the first Exercise Black Skies (EBS) in 2008.

EBS08 had three overarching objectives, which were subsequently captured in the three-part motto: "Prepare, Evaluate, Demonstrate". The first objective was to provide an opportunity for participants to prepare for Exercise Pitch Black (PB); a bi-annual, multi-national, live Air Force training exercise hosted by the RAAF in northern Australia. This was achieved by replicating many of the important characteristics of PB08 (e.g., airspace, order of battle, mission types, unit role assignments) during simulated missions in the DSTO Air Operations Simulation Centre (AOSC). The second objective of EBS08 was to use these realistic mission scenarios to evaluate the tools and methods of SCT in an ecologically valid way. Importantly, EBS08 incorporated a more thorough approach to training effectiveness evaluation than the Pacific Link exercises; including investigation of the transfer of performance benefits from the synthetic exercise to live PB missions. To facilitate these objectives, EBS08 was conducted just prior to PB08. The third objective of EBS08 was to demonstrate the significant potential of SCT for air combat to RAAF. This was achieved primarily by hosting visits to the exercise by those involved in RAAF training, capability development, and leadership. The participants in EBS08 were a RAAF Air Battle Management (ABM) team from the Air Defence Ground Environment (ADGE). Two more EBS activities have been completed since then, each preceding a live PB exercise. Exercise Black Skies 10 (EBS10) included RAAF F/A-18 participants, an ABM team, as well as Army and Special Forces Joint Terminal Attack Controllers (JTACs). The participants in Exercise Black Skies 12 (EBS12) included RAAF ABM teams from both the ground environment and airborne environment (i.e., E-7A Wedgetail).

The success of these activities led to a request in 2010 for DSTO to support RAAF involvement in the United States Air Force (USAF) -led Exercise Coalition Virtual Flag (CVF). At the time of writing, DSTO has successfully hosted the Australian component of three CVF exercises and preparations are nearing completion for a fourth. The apparent benefits of EBS and CVF for RAAF have led to plans to accelerate the development of RAAF's SCT capabilities.

1.2   Evaluation of the Impact of EBS

Central to the Australian Defence Force Training Model are the requirements to: (1) be effective within available resources, and (2) engage in continuous improvement through iterative training analysis, design, execution, and evaluation [7]. Consistent with these requirements, DSTO support to RAAF SCT has focused on the development and evaluation of relatively low-cost simulation systems and the investigation of methods for evaluating training impacts in order to identify opportunities for improvement. To guide this evaluation, theoretical frameworks of training effectiveness and team performance are required.

The most widely cited and applied framework for training effectiveness evaluation is Kirkpatrick's four-level approach [8]. According to Kirkpatrick, training evaluation should take into account data relating to: (1) trainee reactions, (2) learning, (3) behaviour, and (4) results. While Kirkpatrick's four levels of evaluation have been widely applied, this approach has been the subject of criticism [9]. One important criticism is that it fails to take into account the purpose of evaluation; that is, it provides no guidance on how different kinds of data should be used to guide decisions about training development. Such ambiguity can lead to evaluation that is either inefficient, ineffective, or both [10].

Kraiger [10] presented a "Decision-Based" training evaluation framework that addresses this shortcoming by defining three possible purposes for evaluation, namely: (1) to guide decisions about the investment of time and resources into training capabilities or practices, (2) to provide feedback to learners or those involved in training management and delivery, and (3) to market training systems or methods to organisations, departments, or potential participants. Kraiger's framework links these purposes of evaluation with three specific evaluation targets, each of which relates to different characteristics of training. The evaluation targets are: (1) training content and design, (2) changes in learners, and (3) organisational payoffs. Evaluation targets are further broken down into more specific evaluation foci. Important evaluation foci include the utility and relevance of training materials (related to training content and design), the cognitive, affective, and/or behavioural change expected to occur during training (related to changes in learners), and the transfer of training benefits to on-the-job performance (related to organisational payoffs). Finally, the different methods of data collection that can be used to gather evidence for evaluating each target and focus are identified. For some evaluation targets and foci, participant evaluations are appropriate. For others, observation of work samples (e.g., simulations), written tests, or interviews are better suited. Kraiger's framework is depicted in Figure 1, below.
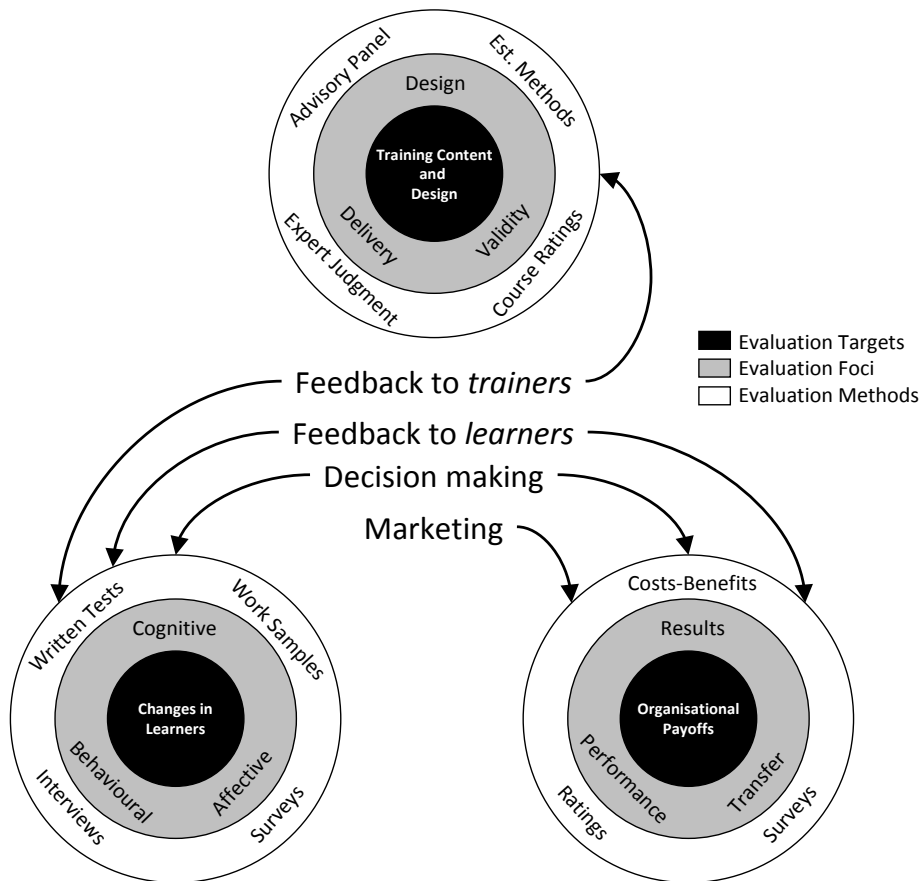

Figure 1. The decision-based training evaluation model presented by Kraiger [10]. An important characteristic of this model is that it explicitly links the targets of training evaluation with evaluation purposes, thereby clarifying how training evaluation data can be used to guide decisions about the development of training tools and methods.

Kraiger's framework makes it clear how evaluation of different aspects of training serves different purposes and involves collecting and analysing different kinds of data. By providing explicit links between targets, foci, methods, and purposes Kraiger's approach closes the loop between training evaluation and training development. It also provides concrete guidance on how the impact of events like EBS should be evaluated.

However, this framework does not fully specify what should be measured. In particular, it is agnostic with regard to the content of particular training domains. This means that in order to properly guide evaluation, it must be considered along with a model of the determinants of performance in the domain of interest.

According to Australian Defence Force (ADF) doctrine, collective training is concerned with the simultaneous and sequential performance of related individual tasks to produce group outcomes [7]. As such, the performance of teams is a central concern. Teams have been defined as sets of two or more people who interact dynamically, interdependently, and adaptively, toward common and valued goals, and who are each assigned specific roles or functions [11]. Both within-team and between-team coordination are likely to be important in air combat. We chose to focus the EBS evaluation on factors operating within teams for two reasons. First, the extensive body of empirical and theoretical research on the determinants of team effectiveness provides a stronger foundation than the relatively modest literature on multi-team systems. Second, it is likely that coordination between teams involves many of the same basic processes as coordination within teams [12, 13, 14].

A framework for understanding the determinants of team effectiveness and the role of SCT in influencing these is presented in Figure 2. This framework is adapted from that which was presented by Kozlowski and Ilgen [15]. It replicates the Input-Process-Output structure and feedback loop of the original. However it also contextualises team effectiveness in SCT, by showing how aspects of training may influence the depicted processes.
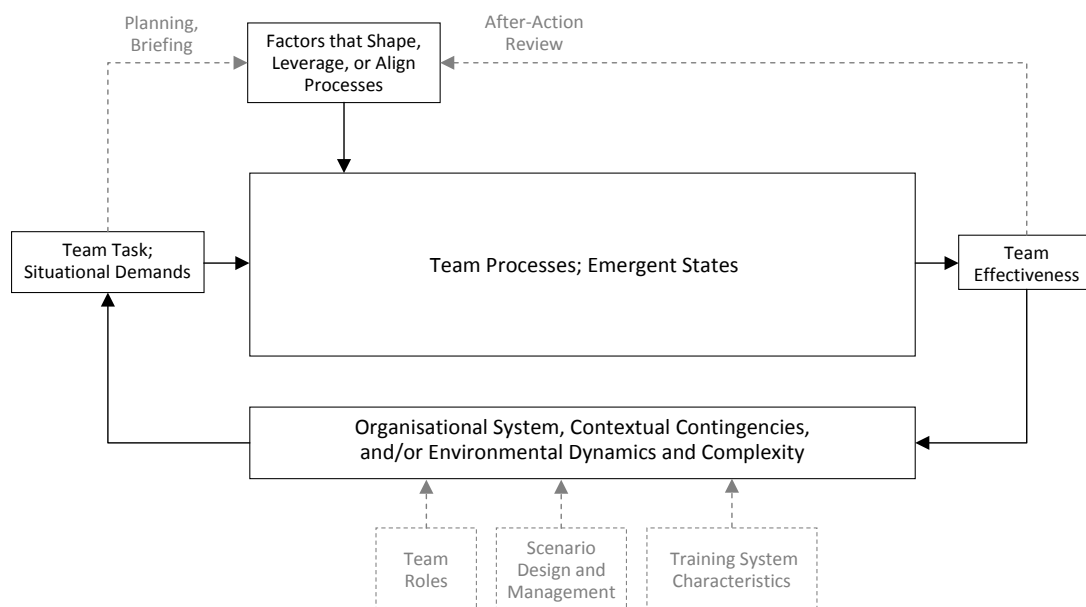

**Figure 2. A conceptual framework of factors underlying team effectiveness. This framework
is based on that which was presented by Kozlowski and Ilgen [15]. It expands on the work
of those authors by showing how aspects of SCT may influence team effectiveness.
Constructs and relations from the original framework are depicted in black text and black
solid lines. Additions are depicted in grey text and grey dashed lines.**

The team effectiveness framework presented in Figure 2 is inherently cyclical. In it, situational characteristics are hypothesised to present teams with demands that they must work to resolve. In the context of SCT, these demands are determined by factors such as the role of the team, the design and management of training scenarios, and the capabilities of training systems. Teams work to resolve situational demands by bringing to bear particular cognitive, affective, and behavioural processes and emergent states (see [15, 16]

for detailed reviews). These factors can also be shaped over time to improve performance. In SCT this typically takes place through processes such as planning, briefing, and after-action review (AAR). When considered in conjunction with Kraiger's [10] framework of training effectiveness evaluation, this description of the factors underpinning team effectiveness has considerable value for guiding an evaluation of the impact of EBS.

1.3   Research Questions

Analyses of the impacts of EBS have included consideration of a wide range of data relating to the factors underlying training and team effectiveness depicted in Figures 1 and 2. A selection of outcomes relating to the behavioural changes in learners as well as organisational payoffs will be described here to provide an overview of some of the more important outcomes obtained thus far. In particular, the data reported below will address the following research questions: Have improvements in team coordination behaviours taken place during EBS? Have participating teams become more effective at achieving their mission objectives during EBS? And, ultimately, have the participants who took part in EBS outperformed those who did not during subsequent, live-flying missions?

2.0   EXERCISE BLACK SKIES PLANNING AND EXECUTION

2.1   Schedule

Each EBS activity followed the same basic schedule. The exercise began on Monday morning and concluded the following Friday afternoon. The Monday of the exercise was either partially (in the case of EBS08) or wholly (in the case of EBS10 and EBS12) devoted to introductory briefings, system familiarisation, and planning activities. Tuesday through Friday consisted of either one (EBS08) or two (EBS10 and 12) vulnerability periods (VULS), during which teams executed their missions. In EBS08 the VUL took place just after lunch each day, with the morning dedicated to planning and briefing. In EBS10 and EBS12, one VUL took place during the morning and one took place during the afternoon. In all cases, VULs lasted for approximately 1.5 hours. Planning processes were abbreviated relative to those that typically take place during live operations in order to provide as many VULs as possible during the week of EBS. The mission scenarios simulated during EBS VULs generally increased in complexity as the week progressed. This increase in complexity was managed in real time by expert White Force (WF) coordinators with the specific aim of presenting the participants with challenging and operationally-realistic problems to solve.

Subject-matter expert (SME) assessors conducted performance evaluations during each EBS VUL. Immediately after the conclusion of each VUL, measurement sessions were conducted during which participants were asked to provide responses to surveys relating to aspects of the systems or activities they had just experienced.

At the conclusion of each measurement session, there was an opportunity for participants to discuss the VUL that had just been completed. In EBS08, this took the form of an in-depth AAR during the afternoon of each day. In EBS10 and 12, this took the form of a relatively quick 'hot wash' following the morning VUL and a more in-depth AAR during the afternoon. These discussions were supported by tools that provided access to 2-D and 3-D replays, operator console screen captures, synchronised recordings of communications from all channels, and ratings and comments by expert assessors.

To achieve the goals of EBS, it was necessary to schedule the exercises just before the live exercise PB. Constraints on participant availability meant that each EBS took place a different period of time before its associated PB. Fortuitously, this enabled an examination of the rate of decay of performance benefits from SCT to live missions (see Section 3.4 below).

2.2   Participants

Six teams of ADF operators (comprised of 22 individuals), have thus far taken part as members of the EBS training audience. This includes four ABM teams (one in EBS08, one in EBS10, two in EBS12, total of 18 individuals) and two CAS teams (each consisting of a JTAC-Pilot pair; both in EBS10).

2.3   Systems

The synthetic environment for EBS consisted of simulator stations for the participants as well as control stations for Red and Blue role-players and the WF exercise coordinators. Each simulation system had its own separate design and implementation with different software and hardware architecture and underlying assumptions. Interoperability was established between the systems using Distributed Interactive Simulation (DIS), Socket-J and Internet Relay Chat (IRC) protocols. Voice communication systems, both radio and intercom networks, were simulated for all participants using a reconfigurable DIS radio application.

ADGE teams engaged in EBS scenarios using simulated workstations that closely replicated the air defence control environment within Australian Regional Operations Centres. Tactical display consoles were implemented using the same Raytheon Solipsys Tactical Display Framework as is used for the real-world ADGE control interface. Each tactical display console presented an air picture that was delivered from a shared Raytheon Solipsys Multi-Source Correlator Tracker (MSCT) system.

F/A-18 Hornet aircrew participated in EBS using the Distributed Air Combat System (DACS) which is a DSTO-developed research simulator configured to model Classic Hornet capabilities. The simulator configuration was tailored for EBS with sub-systems to support CAS scenarios including the LITENING AT FLIR targeting pod and a selection of guided and unguided air-to-ground weapons.

JTAC participants took part using a partial dome display system with a forward field-of-view (200 deg horizontal x 100 deg vertical) looking over the nominated training area. The JTAC was equipped with virtual binoculars, a tripod mounted laser target designator, a DISVOX radio and a joystick-style controller that enabled movement within the virtual environment.

The E-7A Wedgetail mission crew took part in EBS12 using a simulation system that was based upon the Wedgetail Boeing Engineering Test Suite (ETS). The scope of the simulation was to represent the functionality of mission computers, sensors, datalink and display processors. The ETS follows the build of software used within the aircraft and so, from a user interface perspective, the simulation may be considered to be high-fidelity. The simulator was configured with seven operator workstations, although only four mission crew participated in EBS12. Simulated system faults were able to be injected during scenarios to trigger training events. Wedgetail pilot training was not part of the EBS12 exercise and so aircraft flight control was implemented using a Blue Force role-player to command a computer generated entity.

2.4   Measurement

Ratings of the quality of team coordination behaviours and mission effectiveness were obtained during EBS via SME observation and assessment. The quality of ABM team coordination processes was captured using a rating scale based on the Anti-Air Teamwork Observation Measure (ATOM) described by Smith-Jentsch and colleagues [17, 18]. ABM team mission effectiveness was evaluated against a set of mission-specific criteria developed through document review, observation, and consultation with ABM SMEs using the approach to team task analysis described by Annett and his colleagues [19, 20]. Team coordination and effectiveness in CAS was evaluated using observer rating scales based on those developed by Temby, Best, Stephens, and Skinner [21].

2.5 Follow-up Evaluations

After each EBS, follow-up observations were undertaken during the live exercise PB to ascertain the extent to which experiences during synthetic training translated into improved performance during live missions. Ratings of team coordination behaviours and mission effectiveness were obtained using the same instruments that were used during EBS. SME observers evaluated the performance of three ABM teams consisting of members who had previously taken part in EBS (henceforth referred to as 'EBS' teams) as well as that of three ABM teams consisting of members with similar background and experience who had not taken part in EBS (henceforth referred to as 'control' teams).

3.0 RESULTS

3.1 EBS Team Coordination Behaviours

The quality of team coordination behaviours was measured via ratings provided by expert assessors during each EBS VUL. Difference scores were calculated to capture change in ratings of coordination behaviours for each team between the beginning and end of EBS. These differences were then expressed in terms of a percentage of scale maximum score; with positive values indicating performance improvements and negative values indicating performance decrements. Mean scores calculated across six EBS teams are shown in the first column of Figure 3 (error bars represent standard deviations). It can be seen from these data that, on average, there was a gain in performance of around 20% of scale maximum score on these criteria during EBS.

3.2 EBS Team Mission Effectiveness

Team effectiveness was operationalised via expert assessor ratings against role-specific team mission objectives. As was the case for team coordination behaviours, changes in team performance over the course of EBS were evaluated by comparing ratings at the beginning of the activity with those at the end. The second column in Figure 3 shows the average difference score across six EBS teams in terms of a percentage of rating-scale maximum score; with positive values indicating performance improvements and negative values indicating performance decrements (error bars represent standard deviations). It can be seen from these data that, on average, there was a gain in performance of around 10% of scale maximum on these criteria.
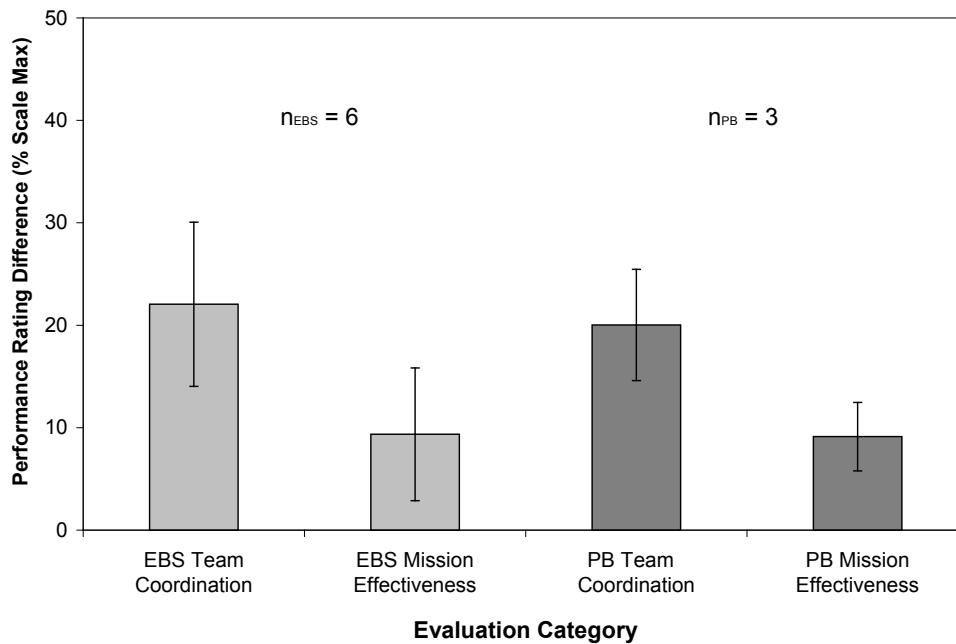
Figure 3. Average EBS and PB difference scores expressed as a percentage of scale maximum. Data from EBS represent the differences in performance ratings of each team between the start and end of the exercise. Data from PB represent the differences in performance ratings between EBS teams and matched control teams during subsequent live missions.

3.3 Performance During Subsequent Live Missions

On-the-job behaviour change is an important indicator of the organisational payoffs from training [10]. After each iteration of EBS, DSTO researchers joined ABM participants during the live exercise PB to observe their performance in relation to that of operators with similar background and experience who had not taken part in EBS. The quality of team coordination behaviours and performance against mission objectives were measured via expert assessor ratings using the same criteria as those used during EBS. Mean difference scores were calculated by comparing the performance ratings of three EBS teams and three matched control teams. These scores are shown in the third and fourth columns of Figure 3 (error bars represent standard deviations). The data shown in Figure 3 indicate that EBS teams outperformed control teams during PB in terms of both team coordination behaviours and mission effectiveness. Interestingly, the magnitude of the performance advantage to EBS teams was similar to the changes in performance observed during EBS – that is, around 20% for coordination behaviours and around 10% for mission effectiveness.

3.4 Decay of Performance Benefits

As described above (see Section 2.1) each iteration of EBS was scheduled a different period of time before its associated live PB exercise due to constraints on the availability of participants. It is reasonable to expect that the performance advantages gained from participation in EBS might decay over time, such that at some point after the synthetic exercise, no advantage would remain for those who had taken part over those who had not. The variability in the timing of EBS and PB provided an opportunity to examine the rate of decay of the benefits obtained by teams who participated. In order to investigate this issue, the difference scores calculated for each EBS team-control team pair were plotted as a function of the time (in days) between the first EBS evaluation and the first PB evaluation. These data are shown in Figure 4 below.
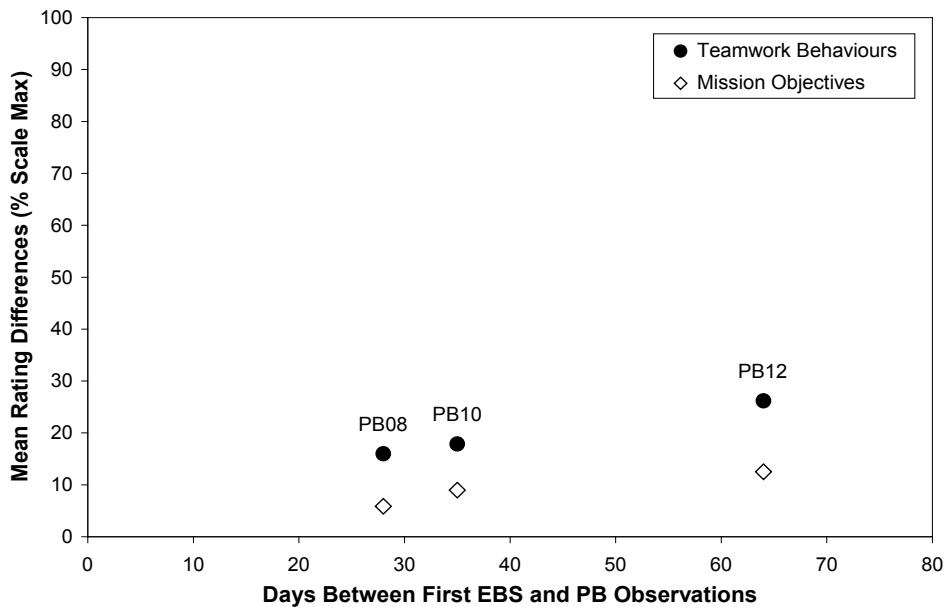
Figure 4. Mean differences between EBS and control teams on assessor ratings of team coordination behaviours (filled circles) and performance against team mission objectives (unfilled diamonds) plotted as a function of the number of days between first EBS observation and first PB observation.

Somewhat surprisingly, there is no evidence from the data in Figure 4 that the performance benefits of taking part in EBS decayed over time: at least during the observed period of around two months. During all three PB exercises, the EBS teams outperformed the control teams by around 10% on mission objectives and by around 20% on team coordination behaviours.

While these data suggest that the benefits of SCT for team coordination and mission effectiveness in the air combat domain may be quite long lasting, there are at least two alternative explanations that must be investigated before this conclusion can be justified. The first is that the data points displayed in Figure 4 do not represent points on a single, relatively flat decay curve, but rather points on three separate decay curves, each associated with the differences between a particular pair of teams. While this is possible, evidence from PB10 suggests that it is unlikely to fully account for the apparent lack of decay. During PB10, the EBS team and control teams were each evaluated several times. In the analyses presented above, scores were averaged across measurement occasions for each team. However, the first and last PB10 evaluations of these teams were conducted on the same days, 10 days apart. This allows an analysis of the rate of decay of performance benefits for those particular teams. On the first occasion that their performance was evaluated during PB10, the EBS team outperformed the control team by an average of around 25% of scale maximum on coordination behaviours and around 11% of scale maximum on mission objectives. On the final occasion that their performance was evaluated (10 days later) the differences between these same two teams on these same criteria were around 29% and 11% of scale maximum respectively. This supports the view that the performance benefits of taking part in EBS were relatively stable for individual teams, at least over a period of several days.

The second alternative explanation for the stability of the effects shown in Figure 4 relates to the measurement procedures used to collect the data. During PB, performance assessments were conducted by expert assessors with experience in the design and management of RAAF ABM training. These assessments were not blinded, meaning that the assessors knew at all times whether they were observing the EBS team or the control team. It is possible that subconscious biases could have influenced these assessments, leading to an overly-positive evaluation of the EBS team relative to that of the control team [22].

However, the practice of real-time, unblinded evaluation by expert assessors is common in training and training research and there are reasons to believe that biases in these ratings can not fully account for these results. For example, empirical evidence presented by Schreiber and his colleagues [23] demonstrated that evaluations provided by unblinded expert assessors in training contexts can be valid. These authors compared the evaluations of blinded and unblinded assessors on a set of F-16 mission scenarios and found similar patterns of performance ratings, despite the fact that blinded assessors did not know whether they were observing performance from the beginning of training or the end of training. Nevertheless, further investigation is required to determine whether rater biases could account for at least some of the apparent stability in the benefits of EBS on subsequent performance during a live exercise and if so, how such effects can be mitigated in the future. These investigations should give consideration to the application of measurement methods other than assessor ratings as well as to methods for effectively blinding assessors.

4.0   CONCLUSION

In order to support the development of RAAF synthetic collective training capabilities, training effectiveness evaluation in EBS has considered a wide range of data sources, each relating to different aspects of training and team effectiveness (see Section 1.2). The data presented here have focused on a subset of these; specifically, data relating to the behavioural changes in learners that have taken place during EBS and the extent to which organisational payoffs have been realised in the form of performance advantages during subsequent live missions. The data presented here suggest that EBS had a positive impact. This impact was realised in terms of positive changes in both team coordination behaviours and performance against mission objectives. Furthermore, these benefits appear to have transferred to subsequent live missions. The apparently long-lasting nature of the benefits is surprising. An understanding of factors affecting the rate of decay of SCT benefits is required to effectively and efficiently integrate SCT with other training activities. This issue will be the subject of further investigation during subsequent EBS activities.

5.0   REFERENCES

[1]   Bennett, W., & Crane, P. (2002). The deliberate application of principles of learning and training strategies within DMT. Proceedings of the NATO Research & Technology Organization Studies, Analysis, and Simulation Panel Conference on Mission Training via Distributed Simulation (SAS-38), Brussels, Belgium, April 2002.

[2]   Schreiber, B.T., & Bennett, W. (2006). Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Summary Report. AFRL-HE-AZ-TR-2006-0015-Vol I. Mesa, Arizona: United States Air Force Research Laboratory.

[3]   Smith, E., & McIntyre, H. (2003). Distributed mission training – How distributed should it be? Proceedings of the NATO Research & Technology Organization, Human Factors and Medicine Panel Symposium on Advanced Technologies for Military Training (RTO-MP-HFM-101), Genoa, Italy, October 2003.

[4]   McIntyre, H.M., & Smith, E. (2013). Key tenets of collective training. In C. Best, G. Galanis, J. Kerry, & R. Sottilare (Eds.), Fundamental Issues in Defense Training and Simulation (pp.125-134). Aldershot, UK: Ashgate.

[5]   Crane, P., Skinner, M., Best, C., Burchat, E., Gehr, S.E., Grabovac, M., Pongracic, H., Robbie, A., & Zamba, M. (2006). Exercise Pacific Link: Coalition distributed mission training using low-cost communications. Proceedings of the 2006 Australasian Simulation Technology and Training Conference (SimTecT), Melbourne, May 2006.

[6]   Best, C., Hasenbosch, S., Skinner, M., Crane, P., Burchat, E., Finch, M., Gehr, S.E., Kam, C., Shanahan, C., & Zamba, M. (2007). Coalition distributed mission training: Exercise Pacific Link 2. Proceedings of the 12th Australian International Aerospace Congress, Melbourne, March 2007.

[7]   Australian Defence Force (2011). Australian Defence Doctrine Publication 7.0 – Training. Canberra: Defence Publishing Service.

[8]   Kirkpatrick, D. (1976). Evaluation of training. In R.L. Craig (Ed.), Training and Development Handbook: A Guide to Human Resource Development (pp. 301-319). New York: McGraw-Hill.

[9]   Holton, E.F. (1996). The flawed four-level evaluation model. Human Resource Development Quarterly, 7(1), 5-21.

[10]   Kraiger, K. (2002). Decision-based evaluation. In K. Kraiger (Ed.), Creating, Implementing, and Managing Effective Training and Development (pp. 331-375). San Francisco: Jossey-Bass.

[11]   Paris, C.R., Salas, E., Cannon-Bowers, J.A. (2000). Teamwork in multi-person systems: A review and analysis. Ergonomics, 43(8), 1052-1075.

[12]   DeChurch, L.A. & Marks, M. (2006). Leadership in multiteam systems. Journal of Applied Psychology, 91(2), 311-329.

[13]   Weaver, S.J., Feitosa, J., Salas, E., Seddon, R., & Vozenilek, J.A. (2013). The theoretical drivers and models of team performance and effectiveness for patient safety. In E. Salas & K. Frush (Eds.), Improving Patient Safety Through Teamwork and Team Training (pp. 3-26). New York: Oxford University Press.

[14]   Drach-Zahavy, A. & Somech, A. (2010). From an intrateam to an interteam perspective of effectiveness: The role of interdependence and boundary activities. Small Group Research, 41(2), 143–174.

[15]   Kozlowski, S.W., & Ilgen, D.R. (2006). Enhancing the effectiveness of work groups and teams. Psychological Science in the Public Interest, 7(3), 77-124.

[16]   Shanahan, C., Best, C., Finch, M., & Sutton, C. (2007). Measurement of the Behavioural, Cognitive, and Motivational Factors Underlying Team Performance. DSTO-RR-0328. Melbourne: Defence Science and Technology Organisation.

[17]   Smith-Jentsch, K. A., Johnston, J. H., & Payne, S. C. (1998). Measuring team related expertise in complex environments. In J.A. Cannon-Bowers & E. Salas (Eds.), Making Decisions Under Stress: Implications for Individual and Team Training (pp. 61-87). Washington: APA.

[18]   Smith-Jentsch, K. A., Zeisig, R., Acton, B., & McPherson, J. (1998). Team Dimensional Training: A strategy for guided team self-correction. In J.A. Cannon-Bowers & E. Salas (Eds.), Making Decisions Under Stress: Implications for Individual and Team Training (pp. 271-297). Washington: APA.

[19]   Annett, J., & Cunningham, D. (2000). Analyzing command team skills. In J.M. Schraagen, S.F. Chipman, & V.L. Shalin (Eds.), Cognitive Task Analysis (pp. 401-415). Mahwah, NJ: LEA.

[20]   Annett, J., Cunningham, D., & Mathias-Jones, P. (2000). A method for measuring team skills. Ergonomics, 43(8), 1076-1094.

[21] Temby, P., Best, C., Stephens, A., Skinner, M. (2009). Tools for the Measurement of Teamwork Factors in Close Air Support. DSTO-TR-2354. Melbourne: Defence Science and Technology Organisation.

[22] Andrews, A., Bowers, C., & Lechin, C. (2013). Factors affecting the reliability and validity of instructor evaluations of trainee competencies. In C. Best, G. Galanis, J. Kerry, and R. Sottilaire (Eds.), Fundamental Issues in Defense Training and Simulation (pp. 275-284). Aldershot, UK: Ashgate.

[23] Schreiber, B.T., Gehr, S.E., & Bennett, W. (2006). Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Real-time and Blind Expert Subjective Assessments of Learning. AFRL-HE-AZ-TR-2006-0015-Vol III. Mesa, Arizona: United States Air Force Research Laboratory.