Chapter 9
Models for and Practice of Continuous Professional Development for Airline Pilots: What We Can Learn from One Regional Airline

Timothy J. Mavin

9.1 Introduction

Aviation is an industry where continuing education and training (CET) is common for all personnel. For some professionals, like airline pilots, CET is not only important, but also mandatory. For example, pilots undergo a regular biannual training and assessment program where they attend (a) classroom-based instruction, and (b) simulator training and assessment sessions. In the classroom, pilots undertake instruction to refresh knowledge of – or be introduced to new – aircraft systems, technical procedures or operational philosophy. In simulator training and assessment sessions, various emergencies will be encountered, requiring pilots to effectively work as a team. In this case, the captain (first in command) assisted by the first officer (second in command) will identify and contain a malfunction and determine the best courses of action. The simulator instructor, known as a *flight examiner* (also *check captain* or *type-rated examiner* in some countries and regions), then makes an assessment of the pilots' individual performance and that of the team. Unfortunately for some pilots, this proficiency examination does not go well, requiring focused retraining followed by a further simulator assessment. In some cases, continued poor performance is career ending.

Over many decades, the aviation industry has come to realize pilot proficiency is a complex interaction of both technical and non-technical skills. How these skills are both developed and assessed in differing training environments – such as classroom and simulator – is a question still under investigation. The purpose of this chapter is to show how two training methods were adapted to the particular needs of an airline through a unique collaboration with a university-based research team. In

T.J. Mavin (✉)
Griffith Institute for Educational Research, Griffith University, Nathan, QLD, Australia
e-mail: t.mavin@griffith.edu.au

particular, I demonstrate (a) how traditional classroom-based instruction moved to a mode that increased the possibility for reflection, and (b) how post-simulator debriefings increase the possibilities of pilots reflecting on their experiences.

9.2 Reflective Practice in Training

An argument premised just over 30 years ago was that traditional training programs do not translate well into the workplace (Schön, 1983). It was proposed that improving life-long learning required a greater focus and analysis on a practitioner's experiences. At the time, reflective practice was formulated as a way for professionals to develop during and after practice. The terms reflection-in-action (reflection during practice) and reflection-on-action (reflection after practice) were coined to assist in making the learning phases of reflective practice more explicit (Schön, 1983, 1987).

Over the decades there have been numerous methods developed to improve reflective practice. One common method used within fields such as health care, human resource development and university education is the reflective journal (O'Connell & Dyment, 2011). The reflective journal requires individuals to write, often informally, about their practice. Although the reflective journal still remains an important method of professional development, it has not been without criticism. For instance, there have been arguments that writing can be superficial (Betts, 2004) or mechanical, and as such, that it does not address the deeper conceptual issues associated with professional practice (Holden & Griggs, 2011). Further critique centres around students not receiving adequate instruction on the purpose or approach to writing, thereby creating some journaling that focuses on perceived instructor expectation rather than reflecting on a writer's experiences and possible learning opportunities (O'Connell & Dyment, 2011).

Other forms of reflective practice include role playing. Here, students' abilities to deal with new situations can be enhanced by role playing possible outcomes prior to real-world practice. For instance, role playing has been used in many professional fields such as health care and aviation; more recently, it has gained increased use for returning soldiers from conflict zones to assist them in dealing with their own past experiences (Hassall & Balfour, in press). In a way, simulation training in aviation allows pilots to role play possible emergencies in high fidelity environments. Here, pilots are able to rehearse flying skills to such a level that they do not need any practice in a real aircraft prior to flying an aircraft with fare-paying customers (Mavin & Murray, 2010).

Critical incident analysis provides another approach to reflective practice. Initially developed in aviation and anesthesia, its main purpose "was on analysing and assessing failures of procedures, or human error, with a view to reducing future risk" (Lister & Crisp, 2007, p. 47). Critical incident analysis is gaining increasing use in the broader health care industry, education and also in social work (Lister &

Crisp, 2007). It has been recognized that critical incident analysis is an effective form of reflection-on-action (Schön, 1983), where participants are able to review previous practice. Furthermore, it has been established that critical incident analysis is helpful for planning future action (Holden & Griggs, 2011), recently referred to as reflection-for-action (Thompson & Pascal, 2012).

As one would expect, increasing use of technology is finding its way into the reflective practice field. Film segments from movies and documentaries become vignettes that facilitate individual and classroom-based discussions about practice. Video recordings taken from actual work practices of students and from professional performances are also used as a means to reflect (Hulsman, Harmsen, & Fabriek, 2009; Todd, 2005). For example, modern aircraft simulators video record cockpit, audio and flight instrument parameters, permit a flight examiner to replay an entire simulator session in the debriefing room; a tool referred to as a debriefing tool.

In regard to the effectiveness of videos, the findings are mixed. Studies show that medical students improve future practice if given opportunities to review their own performance via video recorded reflective sessions (Ward et al., 2003). It has also been demonstrated that improvement can be enhanced if participants are able to conduct reviews whilst accompanied by a senior or more experience person (Lane & Gottlieb, 2004; Scherer, Chang, Meredith, & Battistella, 2003). Yet, larger meta-analyses of studies investigating reflective sessions mediated by personal video fragments (referred to in aviation as a debriefing) indicate they are no more beneficial than sessions that have no video (e.g., Cheng et al., 2014; Tannenbaum & Cerasoli, 2013).

There is increasing literature supporting reflective practice for professional development, though with a caveat of improving its underpinning theory (e.g., Mavin & Roth, 2014a; Thompson & Pascal, 2012) and support via empirical studies (e.g., Koole et al., 2012). In spite of these calls, a key issue continuing to arise is that individuals, asked to reflect on their own practice, must be able to do so. Inherent in this assumption is that once a performance has been completed, individuals reviewing their own performance are capable of appraising that performance. However, considerable literature suggests many poorer performing individuals have great difficulty teasing apart the fundamental strengths and weaknesses of performance (Dunning, Johnson, Ehrlinger, & Kruger, 2003; Dunning & Suls, 2004; Gurung, Daniel, & Landrum, 2012; Sitzmann, Ely, Brown, & Bauer, 2010). That is, when people make incorrect responses, "they are also cursed with an inability to know when their answers, or anyone else's, are right or wrong. They cannot recognize their responses as mistaken, or other people's responses as superior to their own" (Dunning et al., 2003, p. 85). Despite the fact that skilled – or expert – individuals may use reflection as a means of improving performance, it is not known how lower performing individuals – either by virtue of being new to a job or simply through a reduced level of professional development commensurate with (in)experience – gain the greatest benefit from reflective practice, when intrinsically they will have difficulty reflecting. How could a CET program integrate these well-known issues

into new curricula that allow an increase in reflective practice across a range of skilled individuals?

9.3 Background on Pilot Training and Assessment: Decades of Change

In early training, a pilot is familiarized with basic flight instruments, cockpit setup and electrical systems, for instance. Concurrently, pilots learn standard operating procedures on how to operate the aircraft with other crew members, including other pilots, cabin crew, air traffic control, and company personnel (e.g., ground engineers and passenger boarding staff). In early instructional phases, pilots will use their understanding of basic aircraft systems and procedures to conduct training exercises in the simulator, including initial cockpit setup for departure, engine start, and taxiing. As pilots develop greater awareness of systems, simulators are used to integrate this knowledge into the context of both normal and non-normal flight situations. On completion of a training program – generally lasting 6–10 weeks – a majority of pilots develop sufficient skills and proficiency to be accredited to fly a particular aircraft type. Before a pilot is fully endorsed to fly with passengers and other non-training pilots, they undergo an examination in which a flight examiner assesses their skills and performance levels. How might the examiners accomplish their task?

As mentioned, it is mandated that airline pilots must undergo CET and assessment, though these have varied over the years. Studies in the late 1980s and through the 1990s discovered CET and assessment for airline pilots had a technical focus. For example, CET emphasized engineering and systems of the aircraft with extensive assessment of flight manoeuvres (Mavin & Murray, 2010). Even though this approach to technical proficiency remains important (Johnston, Rushby, & Maclean, 2000), it does not fully encompass the reasons why aircraft incidents and accidents were occurring over this period of time. Empirical research demonstrated that whereas technical knowledge of aircraft systems (e.g. aerodynamics, electrics, hydraulics), basic aeronautical knowledge (navigation and rules of the air), and technical skills (manipulation of actual aircraft) are important, these tended not to be the main reasons for the vast majority of aircraft accidents. Instead, skills associated with decision making, teamwork, communication, situational awareness, and management were often identified to have been the root cause of fatal aircraft accidents (Flin, O'Connor, & Crichton, 2008).

As a means of supplementing technical training curricula, airlines developed *crew resource management* material as a way to emphasize effective and efficient teamwork (e.g., Helmreich, Merritt, & Wilhelm, 1999). Though these so-called soft skills – now referred to as non-technical skills – continued to be integral to CET via a variety of didactic methods (e.g., theory concepts and critical incident analysis) there was a level of apprehension within national aviation regulatory authorities that these skills were not assessed in the simulator or aircraft. To effect change, the Joint

Aviation Authorities in Europe developed a separate system to assess non-technical skills; the assessment of technical skills remained unchanged. The non-technical skills system developed was referred to as NOTECHS and consisted of four non-technical skills categories of co-cooperation, leadership and management, situational awareness, and decision making (Flin et al., 2003). To further improve and clarify these categories, each was further divided into sub-categories. For example, *leadership and management* was divided into u*se of authority*, *maintaining standards*, *planning and coordination*, and *workload management* (Flin et al.). To further assist flight examiners in assessing pilot performance, each sub-category was augmented by means of "word pictures" to describe poor and good performance. For example, *use of authority* had word pictures for *poor performance* including "hinders or withholds crew involvement," "passive," and "does not show initiative for decisions," and "own position not recognizable." Good performance on the other hand was described as "takes initiative to ensure crew involvement and task completion", "takes command if situation requires" and "advocates own position" (Flin et al., p. 104). The aim was to make the assessment of non-technical skills mandatory, and to foster implementation by means of newly developed assessment tools.

Over the last decade, NOTECHS (and its many variants developed by airlines) are used to assess pilot performance. Even though there is great support for its use in practice, there has been some questioning of the separation of technical and non-technical skills (Mavin & Roth, 2014b). The main theme of these questions was orientated around the reality of practice. Specifically, when flight examiners assess pilots, do they see a separation of technical and non-technical skills? Furthermore, do flight examiners place a greater emphasis on some skills compared to others, as per the theories of compensatory and non-compensatory skills (e.g., Brannick & Brannick, 1989)?

In response to these concerns, a unified model of performance was developed that did not split performance into technical or non-technical skills. Newer models combine performance dimensions of flying skills and technical knowledge (traditionally technical skills) and situational awareness, decision making, management and communication (non-technical skills) into an integrated model of performance (Mavin, Roth, & Dekker, 2013). Mavin and colleagues also investigated the importance of individual performance dimensions and how they might relate. For example, a pilot during an emergency may become distracted and allow the aircraft to exceed a flight tolerance (sub-category – aircraft flown within tolerances), such as an airspeed or altitude (rather like being distracted while driving a car and running into the curb). However, this may be due to inefficient management of crew tasks (management of crew). In respect to the importance of skills, an aircraft outside of parameters (akin to running off the road) is viewed as a non-negotiable issue. Yet the cause may have been poor management (don't use your mobile phone). Seven airlines in the Tasman and the Australian military use this model for assessment of pilot performance (MAPP) (see Fig. 9.1) as their primary framework for training and evaluation purposes. Here, the MAPP provides a conceptual framework for pilots to assess performance by combining technical and non-technical skills
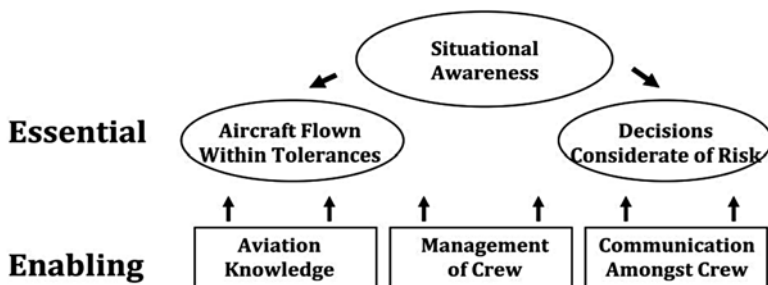
Fig. 9.1 Model for assessing pilots' performance (MAPP) (1,200 dpi)

(traditionally separated) and offers a visual presentation of the hierarchy and causal relationship of skills.

In the following sections, I show how one airline evolved two training modalities. I begin by overviewing the airline's initial crew resource management training, which had an identifiable lack of translating theory into practice, no different from the problem identified decades before in other professions (Schön, 1983, 1987). Through a change in classroom-based curriculum it is illustrated how CET moved from traditional instruction to one based on performance assessment training. That is, I show how making performance increasingly explicit, as per the MAPP, changes the focus of future reflective practice. I then describe and discuss simulator-training principles to show that even with improved understanding of performance, pilots exposed to high workload|stress assessment have difficulty recalling their previous performance. This then entails further refinements to the CET program.

9.4 Changes in Classroom-Based Instruction: Moving Towards Reflective Practice

A decade ago, the airline partner made strategic changes to their CET program in an attempt to come to grips with the findings emanating from the global aviation research community. It was clear that a move towards more non-technical skills content was required if the airline was to acknowledge current worldwide trends in accidents. The fundamental change expressed itself as an increased emphasis on theory, especially skills areas like situational awareness, decision making, management, and communication (Flin et al., 2003; Helmreich et al., 1999). It was assumed that this approach would transfer well to the flight deck of the aircraft. To confirm that the CET program was working, the participating airline developed an assessment instrument for technical and non-technical skills. It encompassed manipulative skills, knowledge of systems and procedures, automated system usage, execution of procedures, communication, workload management, situational
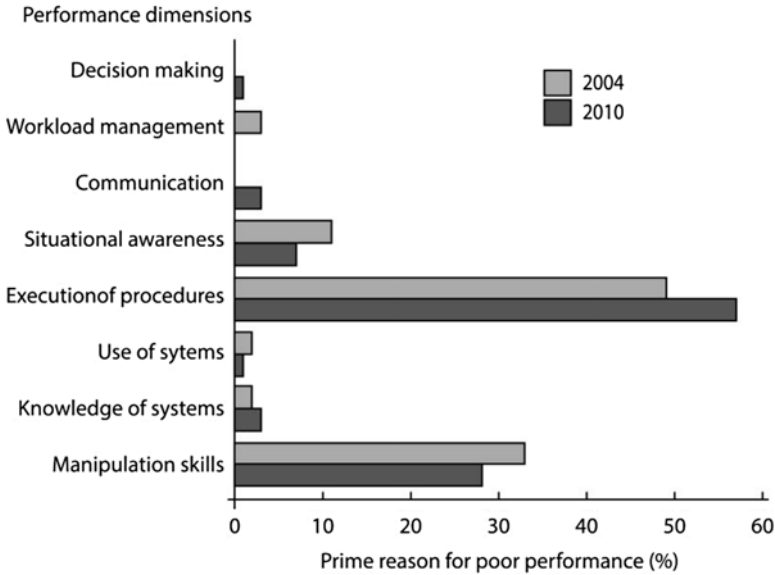
Performance dimensions

Fig. 9.2 A comparison of human factor elements measured in 2004 and 2010 shows little change after 6 years of training investment (1,200 dpi)

awareness, decision making and problem solving; a clear match of areas identified as problematic were taught and assessed.

The new CET program was implemented with baseline measures taken in early 2004 to determine that (a) company pilots were interpreting and correctly using the human factors elements in practice, and (b) flight examiners were assessing the new human factor elements. Another measurement taken in 2010 identified little if any change (see Fig. 9.2). To be more specific, after 6 years of investing in the new CET program, the focus of assessment remained on the technical skills of (a) execution of procedures, and (b) manipulation skills (Munro & Mavin, 2012).

As a result of this second assessment, the airline investigated new approaches to both classroom training and simulator assessment. One of the fundamental changes was the implementation of MAPP as its technical and non-technical skills philosophy. The reasoning for this was that the airline training team identified that the MAPP better represented the way that flight examiners assessed. This decision naturally brought the airline and the university-based research team closer together. In the meetings between airline and researchers soon after these early decisions, it became apparent that current practices were inappropriate: teaching theory and hoping for transfer into the flight deck was not working. Furthermore, it was identified that crew, even after years of additional training, were returning every 6 months to simulator assessments with little if any change.

A new training system was designed using standard instructional design principles including performance objectives, assessment instruments, instructional strategies, and instructional materials. The vision was to teach pilots fundamental skills

Management	1	2	3	4	5
• Workload • Control • Cooperation • Threats & Errors	• Ineffective organisation of crew tasks. • Inability to control self or crew member performance. • Interaction was negligible, or disrupted team effectiveness. • Serious threats or errors not mitigated or managed.	• Inefficient organisation of crew tasks. • Controlled self or crew member actions, though with difficulty. • Interacted with crew member, but provided limited support. • Threats or errors not well mitigated or managed.	• Adequate organisation of crew tasks. • Controlled self or crew members performance; disagreements resolved. • Interacted with crew member. • Most threats managed; most errors trapped.	• Crew member tasks effectively organised. • Effective control of self or crew to achieve expected performance. • Considered other crew to improve team performance. • Threats identified and managed; errors trapped.	• Tasks organised so challenging aspects of flight appeared easy. • Effective control of self or crew, even in a challenging situation. • Interaction with and consideration of crew maximised performance. • TEM well integrated.

Fig. 9.3 The word pictures used to assess management, on a scale from 1 to 5 (1,200 dpi)

of assessment irrespective of their rank. Developing assessment skills within the entire pilot group (including flight examiner, captain and first officer) was thought to improve self-reflective capacities both when pilots were assessed by a flight examiner and also during normal operations when the captain and first officer flew together (Mavin & Roth, 2014a).

The framework for developing assessment instruments came from the performance dimensions contained within the MAPP (e.g., situational awareness, decision making, aircraft flight, technical knowledge, management, and communication). The instrument was in the form of a rubric with word pictures used to describe performance levels from 1 through to 5 (*poor* to *very good* performance). The rationale was that "a holistic rubric is more conducive to providing global judgment of attainment of a benchmark standard at a program level" (Riebe & Jackson, 2014, p. 329). For example, management was a fundamental component of the MAPP; the management word picture described could be graded as a 1 (*poor performance*) all the way through to 5 (*very good performance*) (see Fig. 9.3). Coupled with the use of the MAPP, the assessment instrument enabled finer-grade assessment of each performance dimension with the ability for causal issues coming from the diagrammatic use of the MAPP (see Fig. 9.1). To return to the previous example, the instrument could show that "workload management" was the reason the aircraft was out of tolerance (or you were using your phone). Here, pilots would use maybe a 1 or 2 (see Fig. 9.3) depending on the word picture matching the performance.

As outlined, traditional methods of instruction were not transferring well from the classroom to the flight deck. Given that improving reflective practice was the objective of the training program, it was suggested that training all pilots in the area of assessment, by focusing on a similar method used for inter-rater reliability training, would assist in this area of transfer. Again, the aim was to align how pilots of all ranks assess. Inter-rater reliability training is a technique known for increasing assessment consistency between raters; it is detailed and requires an ongoing commitment (Holt, Hansberger, & Boehm-Davis, 2002). It has three main areas of focus: (a) performance dimension training, (b) behaviour observation training, and (c) frame of reference training.

Performance dimension training familiarizes students with assessment material being used (e.g., Baker & Dismukes, 2002). For example, pilots would be introduced to the MAPP and its fundamentals, followed by the new assessment instrument. This would also incorporate a detailed review of all performance dimensions such as the management field and its 1–5 rating scale (see Fig. 9.3). Behaviour observation training follows, where pilots are taught how to categorize and differentiate each performance dimension, including knowing the difference between *management*

and *communication*. The last step in the training is frame of reference training, where pilots are given the "multidimensionality of performance, defining performance dimensions, [which] provides a sample of behavioral incidents representing each dimension" while also allowing "practice and feedback" (Woehr & Huffcutt, 1994, p. 192). Here, the most effective method has been identified as video assessment. By utilizing appropriate assessment forms, students assess videos and obtain feedback from other pilots and the instructor.

As video assessment was a fundamental instructional strategy in this new training system, the airline invested in developing realistic flight video scenarios. This occurred by filming company pilots in various scenarios in the company simulator. The videotaped scenarios (ranging in length from 2 to 7 min) featured a variety of normal and non-normal situations (non-normal can describe any situation not usually experienced in flight, e.g., sick passenger, hydraulic failure, or engine fire) in fine and severe weather conditions. It was understood that developing realistic training materials would link directly to the pilots' world.

Even though performance assessment training was normally limited to flight examiners, the new aim for classroom-based instruction was to improve the assessment skills of all pilots. It was argued that having pilots assess the performance of peers in the scenarios – using the performance dimensions from the MAPP and the new assessment instrument – would create a stronger and more authentic link between theory and everyday work in the cockpit. Furthermore, as the MAPP provided a conceptual model of how a flight examiner assesses performance, junior pilots and those pilots who had performance issues would better understand the reasons why pilots in a video vignette may have failed, and the probable causal factors.

9.4.1   A Typical Training Day

After initial introductions, a typical training day in the revised classroom began with the projection of a specific scenario: in this instance, where two pilots taxi an aircraft in poor visibility to the runway. When the cabin crew (flight attendant) calls to announce that a passenger is very sick, the pilots are distracted. As the situation unfolds, they eventually find themselves on the wrong taxiway.

After playing this clip, the instructor asked workshop participants to individually assess each pilot in the scenario. Participants were specifically directed to comment on the nature of the problems that existed with the pilot's performance. In other words, what was being identified through the participants' eyes as the key reason that made them either concerned or pleased with the observed performance? Participants also were asked to identify how the problem could be fixed. The solution was to be stated in terms of the view a flight examiner would take: "What would the flight examiner emphasize during a (possible) debriefing that would enable the (scenario) pilots to improve their performance?" The last component was the actual debriefing: How would they go about discussing this issue with the pilots? That is,

all pilots were being asked to role play a flight examiner. The lesson focused on a sequence: problem → fix → debrief. After rating the pilots individually, participants teamed up with a peer. Finally, they discussed previous findings in groups of four. With a class of eight, this process would take 1-h prior to the instructor bringing the groups to a general discussion. No assessment tools were provided; pilots were required to make judgments using only their own experiences to this point.

What is surprising using this method is the difficulties pilots experience in identifying a consistent approach to the problem and the causal reasoning (fix) why the aircraft ended up on the wrong taxiway. Over a period of years, the airline had developed a culture of focusing on technical skills (specifically, execution of procedures and manipulation skills) even though a broad spectrum of technical and non-technical skills were taught and assessed (see Fig. 9.2). Accordingly, the general aspects that workshop participants identified as problematic related to failure in following procedures, communicative trouble, not stopping the aircraft when they first received the call from the cabin crew, crew inefficiency (e.g., the first officer was hopeless, the captain was hopeless), management, the aircraft being on the wrong taxiway, loss of situational awareness, and so on. During the subsequent classroom discussion, the instructor listed on the white board, under the headings *problem*, *fix*, and *debrief*, all items that had been identified by the pilots. On most occasions, there were over 30 issues produced for the 4-min clip. Fundamentally, there was no common framework among the pilot group for assessing performance.

After students had conducted their first individual assessment of a video and as a group collated scores and reasons on the white board, the MAPP was introduced. This demonstrated to the pilots in the classroom that the pilots in the video had experienced a reduction in essential skills (see Fig. 9.1) of situational awareness, which led to a failure to maintain the aircraft within tolerance (the aircraft was taxied in contravention of actual clearance). However, as use of the MAPP could illustrate, the fundamental reason that the aircraft was in this position was the failure of the crew to *manage* the incident. It would therefore follow that the debriefing meeting was to focus on management skills. For some workshop participants, there was a little confusion on why *stopping* was not the way to fix the problem. However, stopping the aircraft would have been an *event fix* rather than giving pilots *broader skills* that could be transferred to other events.

When the MAPP had been introduced and discussed, the assessment instrument was then introduced to the pilots. The workshop leaders demonstrated that management was the prime causal factor for poor performance. With the assessment instrument it was categorized under "workload management" with a rating of 1 (*Ineffective organization of crew tasks*) (see Fig. 9.3). On completion on this first video, theory pertaining to each performance dimension – now no different from traditional training – was introduced, using the first video as a frame of reference or anchor.

9.4.2   Summary

What has been described here is a new approach to how theoretical training is now conducted within this airline. Rather than theory being taught in the classroom first, videos have become the central focus of training. Pilots are required to use previous experience in an attempt to identify strengths and weaknesses in the vignettes viewed. The instructor then facilitates discussion among students about their interpretations of the performance. The ongoing research shows that many pilots now are better in identifying strengths and weaknesses of their performances, thus improving reflective abilities.

9.5   Simulation Instruction: Getting Pilots to Remember What Actually Happened

Simulation for both CET and assessment is fundamental to aviation today. In their most sophisticated form, simulators used for the last 20–30 years have been able to replicate the fidelity of a modern airliner in almost every way (Mavin & Murray, 2010). This level of sophistication now allows pilots to make their first real-aircraft training flight with passengers. The use of simulation within airlines has two prime purposes. The first is *type rating* training, where pilots initially learn to fly a particular aircraft type. The type rating includes classroom-based and simulation training lasting approximately 6 weeks prior to pilots undergoing flight training in a real aircraft. The second use of the simulator is for quality assurance. It is a requirement that airline pilots undergo a regular training and assessment program. Every 6 months (slight differences do occur between countries and airlines) all airline pilots undergo a 2-day training and assessment program. Each day consists of a 6-h training footprint that includes a 1-h briefing and 4-h simulator session, concluding with a 1-h debriefing. Even though most airlines encourage training as an underlying philosophy for each simulator session, it is still incumbent on pilots to attain proficiency by the end of the simulator session.

The simulator session can encompass a variety of training and assessment tasks. For example, when pilots first enter a simulator after the briefing, they will generally spend approximately 10 min setting up for departure. Once the setup is completed, the flight examiner (sitting at an operator console in the rear of the simulator) directs the session, operating the simulator and acting on behalf of traffic control, ground engineer, cabin crew and other aircraft.

A couple of specific sessions are always conducted. The first is a manoeuvre-based sequence. Here, the flight examiner sets up a specific manoeuvre for the pilots to conduct. In early training it could be as simple as an engine start or a rejected takeoff, or a more complex manoeuvre where an engine fails as the aircraft is rotating during takeoff. The pilots will be required to fly the aircraft to a safe altitude, secure the engine (putting out an engine fire in some cases) and land the aircraft at

a suitable airport, which could be the airport of departure or another airport depending on weather conditions. Generally in manoeuvre-based sequences, the aircraft is continually repositioned to allow for the next manoeuvre, which in itself creates issues for some pilots; these are discussed later.

The second type of training is *line-oriented flight training*. Here a normal flight is planned, with pilots during the briefing generally spending time planning for the flight. They are provided with normal flight plans, aircraft status, passenger loads, and specific weather, from which information they plan the flight. The aim is for the assessment to be as realistic as possible, with the flight examiner acting only as a traffic controller, a ground engineer, or cabin crew. During the flight, the flight examiner instigates specific non-normal events that the pilots are required to deal with. The events can be as simple as increasingly poor weather at the destination or malfunctions of a single system, or more complicated emergencies where one system malfunction can affect another system.

On completion of the 4-h session, pilots leave the simulator and return to a room to conduct a debriefing with the flight examiner. It is here that the flight examiner uses a variety of artifacts to conduct the debriefing of the crew performance. These artifacts can include notes made by the flight examiner in the simulator, charts and documents used by the pilots during flight, or whiteboard drawings produced on the spot to explain the flight manoeuvres. In some simulators, the debriefing tool (outlined previously) is used to replay selected scenes in the debriefing room. This allows the pilots to look at actual performance from a third-person perspective.

During the time that the university-based research team worked with the airline, we had begun with the assumption that pilots trained in performance assessment would be far better equipped to self-assess in the debriefing. To test our initial hypothesis, our research team began a large study investigating the actual practice of debriefing. In this study, we videotaped 29 entire debriefing sessions. To compare our partner airline with the practices of other airlines, five airlines participated in this study.

The study identified a number of important issues, some of which are outlined here. First, some pilots were emerging from the 4-h simulator session disorientated, especially after manoeuvre-based sequences. Second, all pilots were fatigued, regardless of their performance. That is to say, even pilots who had performed at an exceptional level appeared to be as tired as those pilots who had performed poorly. For example, when asked to review performance in the simulator, pilots had difficulty remembering the sequence in its entirety, thus making reflection difficult. Third, flight examiners were not taking into account the difficulties pilots were having in remembering what actually occurred in the simulator session. On numerous occasions pilots had to clarify what scenario the flight examiner was discussing. Basically, flight examiners without this realization were analyzing and critiquing a specific event; the pilots on the other hand were (a) trying to determine which event the examiner was talking about, and (b) reconstructing what had happened. Finally, flight examiners were not giving pilots ample time during discussions or when answering questions. In other words, examiners did not give enough time after a question was asked, or after an answer, a period called "wait time" (Rowe, 1986).

Fig. 9.4 New debriefing format depicting specific phases (300 dpi)

From this study arose two key questions: Were current debriefings effective, and if debriefings were not effective, how could the practice be changed?

By integrating previous studies the university had conducted with the airline (e.g., Mavin et al., 2013; Roth & Mavin, 2015), other studies on performance assessment (e.g., Dunning & Suls, 2004) and the current debriefing studies, we identified key issues that were making debriefings, and therefore reflection, less effective than they might be. These included (a) pilots of different rank assessing performance differently, (b) a disparity between ability to perform and ability to self-assess, (c) pilots being fatigued (no matter performance level), (d) some pilots being disorientated by numerous simulator repositions, (e) flight examiners doing most of the talking, (f) pilots finding it difficult to remember, and (g) flight examiners exhibiting poor wait time during discussion. To address these issues a new framework for debriefing was developed. It consisted of five phases, as shown in Fig. 9.4. Here, pilots initially review plan for simulator session (Phase 1); review positive and negative performance events (Phase 2); review a selected performance event in detail by talking through event, reviewing the simulator video of the event and assessing the performance by MAPP (Phases 3 and 4); and then final review (Phase 5).

In the first phase of this new framework, flight examiners encouraged pilots (rather than directing them) to provide an overview of simulator session details, simply by asking, "What were we planning to do?" This enabled pilots, especially those who were disorientated, to develop a clear understanding of what was meant to occur in the simulator. It also allowed the pilots time to talk more, with wait time being an important skill now learned by flight examiners. Surprisingly, this stage, which had usually not been occurring previously, was now taking as long as 12 min.

The second phase required pilots to identify positive and negative performance areas. We had identified, as had other studies, that pilots of different rank (i.e. flight examiner, captain and first officer) assessed performance differently (e.g., Mavin et al., 2013). At this stage the flight examiners were encouraged to be noncommittal in their interpretations of the pilots' perceptions of their own performance. As part of the third phase, flight examiners and pilots identified a particular scenario identified as either well done or in need of improvement. Flight examiners would then encourage pilots to relive (remember) that experience, or what we referred to as the *first-person* experience. This reliving of the experience was important, as it had been identified that pilots were having problems trying to do so. The flight examiner encouraged the pilot to describe (a) what they were doing, and (b) what they were

thinking. In cases when the airline had access to a debriefing tool, the flight examiner replayed the scene to the pilots. This is what I called the *third-person* perspective. It was only after these reliving experiences, both from a first- and third-person perspective, that the flight examiners engaged in analysis of performance. On completion of the analysis, another scenario would be selected for the pilots to relive.

Depending on time available, the final phase required the flight examiner to summarise the simulator session, or what we refer to as main learning points. As can be seen in Fig. 9.4, debriefing is initially linear (Phases 1 and 2) leading to a cyclical review where a number of focus areas are covered (Phases 3 and 4) followed with a final review (Phase 5). Comments from some pilots with whom we discussed this process noted it "is far better." There is now "a definite change in philosophy, how the debrief was run, very handy for us, for me anyway, use the time to chronologically list as a team, because you can't remember it … often after a simulator you're tired, you learn in the simulator and you learn in the debrief."

9.5.1   Summary

I initially made the assumption that pilots, previously trained in performance assessment, would be able to correctly evaluate their own performance on completion of a simulator session. However our debriefing study demonstrated that for a pilot to be able to assess performance they must first *make present again* what had gone before, prior to being able to reflect. This makes sense, as many studies demonstrate that cognition within the flight deck of an aircraft is situated and distributed (e.g., Henriqson, van Winsen, Saurin, & Dekker, 2011; Hutchins, 1995; Roth, Mavin, & Munro, 2014). That is, past experience of performance is not contained within a single person: in fact it is spread across the captain, first officer and aircraft systems, thus requiring a process – Phase 1 through 3 – to bring the past to the present. Because pilots can talk between them to reconstruct the simulator session, hear and see each other in the video, important aspects of the flight become present again. This is enhanced by the representation of instruments, which show exactly what pilots had available in the simulator. The third-person view, and the recalled first-person experience, increase the quantity and quality of represented experience, which then was available for analysis, assessment (using the MAPP and assessment instrument), and learning.

9.6   Classroom and Simulation Training: Improving Learning

In the foregoing sections, I described the ways in which an airline had changed two CET training programs it had been using: classroom-based instruction and simulator training associated with assessment. In the early years the airline increased the

focus on training of non-technical skills due to evidence being presented from the world aviation community (e.g., Helmreich et al., 1999). However, teaching non-technical skills theory (Flin et al., 2003) in the class did not transfer well into practice.

As part of a university|airline collaboration we had identified a couple of key issues. First, the best learning requires socially and physically authentic environments (e.g., Ericsson, 2008). Classroom-based training, by virtue of its decontexualized setting, can make authentic instruction difficult (Roth, 1995). Nevertheless, our use of videos – thereby consistent with existing research (e.g., Merriam, Caffarella, & Baumgartner, 2007) – provided an increased level of authenticity. The videos afforded anchored instruction (Merriam et al., 2007), whereby practice (the flight deck) was brought into the classroom. Using this approach allowed us to integrate theory with the practice of flying in the way that the pilots are familiar. Secondly, there was a consistent message that newer or poorer performing pilots had difficulty understanding assessment decisions of the more experienced flight examiners. This was consistent with other studies into performance assessment (e.g., Dunning & Suls, 2004; Dunning et al., 2003; Gurung et al., 2012; Sitzmann et al., 2010) and our study with airline pilots (Mavin et al., 2013). The focus of training therefore moved towards performance assessment training as a means of improving all pilots' ability to reflect on practice.

Yet there still remains a need to step beyond looking at pilots merely as individuals and to take a broader view on performance, which can only be learnt when the pilots are viewed within actual or simulated work settings. In simulation training, it was assumed that pilots now trained in performance assessment would be better equipped to assess their own performance. Given that pilots are trained to assess performance, and given that they had just completed the events in the last couple of hours, it was assumed that reflecting back would be a simple affair. Again, our research suggests this is simply not the case. We discovered that while a flight examiner can quite accurately identify events to be discussed, pilots under assessment actually struggle to recall what occurred during their simulator exercises. In the revised training program, there now existed an interim step between reflection on practice: an emphasis on recalling or re-remembering events that they had conducted. By changing flight examiners' debriefing framework, as illustrated in Fig. 9.4, we assisted pilots in recalling events prior to the analysis of performance via assessment instruments, thus allowing pilots to better use the assessment skills taught in the classroom, and thereby improving reflection.

9.6.1 Broader Implications for CET

Our work has a number of important implications that may assist other professional fields. The first implication for CET concerns an individual's ability to recall past events. Here there are two issues that are at play. The first is associated with memory, specifically explicit and implicit memory. Explicit memory is viewed as the

purposeful referencing of prior experiences, and is linked to conscious recollection (McKone & French, 2001). On the other hand implicit memory often requires little if any conscious effort to recall; implicit memory proficiency often relates to exposure to previous tasks, and "makes no reference back to any particular encounter with the stimulus" (McKone & French, p. 806).

For many years explicit memory has been categorized to include areas like episodic (past events) and semantics (factual information) memory. Of late these categories have been subject to increasing debate, with some research questioning its separation from implicit memory (Goujon, Didierjean, & Poulet, 2014), and others stating that explicit memory exists only as mental representations, and is thus unable to be observed (Grimm, 2014). Yet, while we improve our understanding of how knowledge may be categorized, we are still aware that when we recall past experiences – specifically explicit memory – recollections can be inchoate or unclear (Roth & Jornet, 2013). For example, when we ask individuals to recall a performance in a simulator, classroom or even an operating theater, the recollection may be unclear, incomplete and in some cases provide only what was done, not what an individual was thinking, or vice versa. It requires increased effort on the part of the individual and trainer to elicit a detailed view of past experience.

With implicit memory on the other hand, we have little if any recollection of how actions come to play, are learnt or acquired. Nevertheless when asked to perform a task we are familiar with, like playing piano or operating on a patient, we are able to perform these familiar tasks flawlessly (Roediger, 1990). Here implicit representations are difficult, as they can be viewed as sensory-motor or embodied (e.g., Sheets-Johnstone, 2011). For instance, when asking an individual to describe an action, they may have difficulty recalling what they did, how they performed a task, and in some cases may even require acting out the task to assist in the recall.

The second issue to be discussed regarding memory relates to *distributed cognition*. Distributed cognition describes how, within a group such as a team of professionals, cognition is not held internally by one individual, but shared across multiple team members, and aided by the use of artifacts (Hutchins, 1995). Here the argument is that within teams the "operation of a distributed-cognitive system is parallel in that multiple people and artifacts work simultaneously" (Cheon, 2014). To be precise, multifaceted work environments are viewed as a complex socio-technical system, like an intensive care unit (ICU) at a hospital (e.g. Rajkomar & Blandford, 2012). For example, when a doctor in hospital examines a patient, they make a diagnosis and prescribe treatments to a patient – be it physical therapy or pharmaceutical drugs. A nurse will action these treatments or prescriptions (by referencing to patient records) while at the same time caring for a patient's daily needs, such as measuring vital signs, and again recording these on a patient's chart. A doctor, be it the same one or another, will return again to review a patient, and record their progress, and so on. What is known about the patient within this clinical setting is not contained within one individual, but distributed amongst doctors, nurses and artifacts. In this case, remembering all there is about the patient would require the team of doctors and nurses to discuss the patient in vivid terms, while referencing arti-

facts such as patient records. This would allow for an entire picture of the patient's well-being, health and care to be clearly developed and understood.

The second implication for CET is not in focusing on teaching reflection, but rather teaching how to assess performance. Some approaches to CET include reflecting on performance with journals (e.g. O'Connell & Dyment, 2011), critical incident analysis (Lister & Crisp, 2007), reflection on performance aided with the use of videos (e.g. Hulsman et al., 2009; Todd, 2005) or even using instruments that measure the effectiveness of reflection itself (e.g. Thorsen & DeVore, 2013). However, the literature fails to demonstrate the link between an individual's performance and the effectiveness of measuring that performance. As Dewey wrote over a century ago, "To find out what facts, just as they stand, mean, is the object of all discovery; to find out what facts will carry out, substantiate, support a given meaning, is the object of all testing" (Dewey, 2012, p. 116). Here Dewey recognizes that meaning is fundamental. Yet, as there appears to be a connection between performance ability and reflective ability – sometimes referred to as a *double paradox* (Dunning et al., 2003) or *ability effect* (Cassidy, 2007) – the research suggests that deliberate approaches to specifically teaching the *meaning* of performance would result in improving reflective ability.

In summary, during CET, the process of learning new procedures – during normal practice or even during times of high workload or stressful events – may require an individual to spend time recollecting and reviewing performance, sometimes with other team members. This process must occur prior to an individual or team being able to assess on action. Of equal importance is the fact that, once a picture has been clearly painted of a performance, it must be determined whether the individual will be able to reflect on their actions in order to assess performance and develop strategies that will realise improvements in their actions.

As a final note, it is worth observing that non-technical skills mentioned in this chapter are not unique to airline pilots. Numerous professions, such as health care, armed services, police, mining and rail, must attend to and contend with non-technical skills (e.g., Flin et al., 2008). While aviation has undoubtedly led the world in non-technical skills training, lessons from our program could well be applicable to other professions.

9.7   Conclusion

The aviation industry is one where CET and assessment are integral to practice. Traditional CET and assessment programs for airline pilots have focused on technical proficiency. It has only been over recent decades that increasing evidence has demonstrated that technical skills deficiencies, while important, are not the full reason for aircraft incidents (Helmreich et al., 1999). The identification of non-technical skills as causal in many accidents has created a change in the direction for pilot training.

The use of classroom-based instruction is necessary to almost all training environments; as would be expected, classroom-based training became the focus for teaching non-technical skills. Nevertheless, as I demonstrate in this chapter, this approach has not worked well for one airline.

The collaborative airline/university research undertaken in this chapter predicted that focusing on performance assessment training in the classroom, for pilots of all ranks, would achieve a greater understanding of non-technical skills. While this approach was reasonable, it revealed only half the story. As evidence mounted identifying the difficulties pilots were having remembering the intense simulator sessions of only 4 h prior, it became apparent that new approaches to debriefing were required. Once the importance of remembering prior to reflection informed the debriefing sessions, the CET program's focus became apparent.

References

Baker, D. P., & Dismukes, R. K. (2002). A framework for understanding crew performance assessment issues. *The International Journal of Aviation Psychology, 12*, 205–222.

Betts, J. (2004). Theology, therapy or picket line? What's the good of reflective practice in management education? *Reflective Practice, 5*, 239–251.

Brannick, M. T., & Brannick, J. P. (1989). Nonlinear and noncompensatory processes in performance evaluation. *Organizational Behavior and Human Decision Processes, 44*, 97–122.

Cassidy, S. (2007). Assessing 'inexperienced' students' ability to self assess: Exploring links with learning style and academic personal control. *Assessment & Evaluation in Higher Education, 32*, 313–330.

Cheng, A., Eppich, W., Grant, V., Sherbino, V., Zendejas, B., & Cook, D. A. (2014). Debriefing for technology-enhanced simulation: A systematic review and meta-analysis. *Medical Education, 48*, 657–666.

Cheon, H. (2014). Distributed cognition in scientific contexts. *Journal for General Philosophy of Science, 45*, 23–33.

Dewey, J. (2012). *How we think*. New Orleans, LA: Quid Pro Books (Original work published 1910).

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*, 83–87.

Dunning, D., & Suls, J. M. (2004). Flawed self-assessment implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*, 69–106.

Ericsson, A. K. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine, 15*, 988–994.

Flin, R., Martin, L., Goeters, K., Hörmann, H., Amalberti, R., Valot, C., et al. (2003). Development of the NOTECHS (non-technical skills) system for assessing pilots' skills. *Human Factors and Aerospace Safety, 3*, 97–119.

Flin, R., O'Connor, P., & Crichton, M. (2008). *Safety at the sharp end*. Aldershot, UK: Ashgate.

Goujon, A., Didierjean, A., & Poulet, S. (2014). The emergence of explicit knowledge from implicit learning. *Memory & Cognition, 42*, 225–236.

Grimm, L. R. (2014). Psychology of knowledge representation. *Wiley Interdisciplinary Reviews. Cognitive Science, 5*, 261–270. doi: 10.1002/wcs.1284.

Gurung, R. A. R., Daniel, D. B., & Landrum, R. E. (2012). A multisite study of learning in introductory psychology courses. *Teaching of Psychology, 39*, 170–175.

Hassall, L., & Balfour, M. (in press). Transitioning home: Research-based theatre with returning servicemen and their families. In B. George & L. Graham Lea (Eds.), *Research-based theatre*. Bristol, UK: Intellect.

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*, 19–32.

Henriqson, E., van Winsen, R., Saurin, T. A., & Dekker, S. W. A. (2011). How a cockpit calculates its speeds and why errors while doing this are so hard to detect. *Cognition, Technology, and Work, 13*, 217–231.

Holden, R., & Griggs, V. (2011). Not more learning logs! A research based perspective on teaching reflective learning within HR professional education. *Human Resource Development International, 14*, 483–491.

Holt, R. W., Hansberger, J. T., & Boehm-Davis, D. A. (2002). Improving rater calibration in aviation: A case study. *The International Journal of Aviation Psychology, 12*, 305–330.

Hulsman, R. L., Harmsen, A. B., & Fabriek, M. (2009). Reflective teaching of medical communication skills with DiViDU: Assessing the level of student reflection on recorded consultations with simulated patients. *Patient Education and Counseling, 74*, 142–149.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science, 19*, 265–288.

Johnston, A. N., Rushby, N., & Maclean, I. (2000). An assistant for crew performance assessment. *The International Journal of Aviation Psychology, 10*, 99–108.

Koole, S., Dornan, T., Aper, L., De Wever, B., Scherpbier, A., Valcke, M., … Derese, A. (2012). Using video-cases to assess student reflection: Development and validation of an instrument. *BMC Medical Education, 12*, 22–29.

Lane, J. L., & Gottlieb, R. P. (2004). Improving the interviewing and self-assessment skills of medical students: Is it time to readopt videotaping as an educational tool? *Ambulatory Paediatrics, 4*, 244–248.

Lister, P. G., & Crisp, B. R. (2007). Critical incident analyses: A practice learning tool for students and practitioners. *Practice: Social Work in Action, 19*, 47–60.

Mavin, T. J., & Murray, P. (2010). The development of airline pilot skills through practice. In S. Billett (Ed.), *Learning through practice: Models, traditions, orientations and approaches* (pp. 268–286). Dordrecht, The Netherlands: Springer.

Mavin, T. J., & Roth, W.-M. (2014a). Between reflection on practice and the practice of reflection: A case study from aviation. *Reflective Practice: International and Multidisciplinary Perspectives, 15*, 651–665.

Mavin, T. J., & Roth, W.-M. (2014b). A holistic view of cockpit performance: An analysis of the assessment discourse of flight examiners. *The International Journal of Aviation Psychology, 24*, 210–227.

Mavin, T. J., Roth, W.-M., & Dekker, S. (2013). Understanding variance in pilot performance ratings: Two studies of flight examiners, captains and first officers assessing the performance of peers. *Aviation Psychology and Applied Human Factors, 3*, 53–62.

McKone, E., & French, B. (2001). In what sense is implicit memory "episodic"? The effect of reinstating environmental context. *Psychonomic Bulletin & Review, 8*, 806–811.

Merriam, S. B., Caffarella, R. S., & Baumgartner, L. M. (2007). *Learning in adulthood: A comprehensive guide*. San Francisco, CA: Wiley.

Munro, I., & Mavin, T. J. (2012, November). *Crawl-walk-run*. In Proceedings of the 10th international symposium of the Australian Aviation Psychology Association. Sydney, Australia: AAvPA.

O'Connell, T. S., & Dyment, J. E. (2011). The case of reflective journals: Is the jury still out? *Reflective Practice: International and Multidisciplinary Perspectives, 12*, 47–59.

Rajkomar, A., & Blandford, A. (2012). Understanding infusion administration in the ICU through distributed cognition. *Journal of Biomedical Informatics, 45*, 580–590.

Riebe, L., & Jackson, D. (2014). The use of rubrics in benchmarking and assessing employability skills. *Journal of Management Education, 38*, 319–344.

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45*(9), 1043–1056.

Roth, W.-M. (1995). *Authentic school science*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Roth, W.-M., & Jornet, A. (2013). Toward a theory of experience. *Science Education, 98*, 106–126.

Roth, W.-M., & Mavin, T. J. (2015). Peer assessment of aviation performance: Inconsistent for good reasons. *Cognitive Science*. doi:10.1111/cogs.12152.

Roth, W.-M., Mavin, T. J., Munro, I. (2014). How a cockpit forgets speeds (and speed-related events): Toward a kinetic description of joint cognitive systems. *Cognition, Technology and Work 2*, 1–21.

Rowe, M. B. (1986). Wait time: Slowing down may be a way of speeding up! *Journal of Teacher Education, 37*, 43–50.

Scherer, L. A., Chang, M. C., Meredith, J. W., & Battistella, F. D. (2003). Videotape review leads to rapid and sustained learning. *American Journal of Surgery, 185*, 516–520.

Schön, D. A. (1983). *The reflective practitioner*. New York, NY: Basic Books.

Schön, D. A. (1987). *Educating the reflective practitioner*. San Francisco, CA: Jossey-Bass.

Sheets-Johnstone, M. (2011). *The primacy of movement*. [DX Reader version]. Retrieved from http://librarycatalogue.griffith.edu.au/record=b1932512.

Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education, 9*, 169–191.

Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do team and individual debriefs enhance performance? A meta-analysis. *Human Factors, 55*, 231–245.

Thompson, N., & Pascal, J. (2012). Developing critically reflective practice. *Reflective Practice: International and Multidisciplinary Perspectives, 13*, 311–325.

Thorsen, C. A., & DeVore, S. (2013). Analyzing reflection on/for action: A new approach. *Reflective Practice, 14*, 88–103.

Todd, G. (2005). Reflective practice and Socratic dialogue. In C. Johns & D. Freshwater (Eds.), *Transforming nursing through reflective practice* (pp. 38–54). Oxford, UK: Blackwell.

Ward, M., MacRae, H., Schlachta, C., Mamazz, J., Poulin, E., Reznick, R., et al. (2003). Resident self-assessment of operative performance. *American Journal of Surgery, 185*, 521–524.

Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 67*, 189–205.