

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

A holistic view of cockpit performance: An analysis of the assessment discourse of flight examiners

Timothy J Mavin ^{a,*}, Wolff-Michael Roth ^b

^a *Griffith Institute for Educational Research, Building M10, Griffith University, Mt Gravatt QLD 4122, Australia*

^b *Faculty of Education, MacLaurin Building A557, University of Victoria, Victoria, BC, V8W 3N4, Canada*

Abstract

What pilots do on the job frequently is analyzed in terms of individual skills and human factors. Performances often do not consist of separable skills but of a holistic event, which can be analyzed into irreducible, mutually constitutive moments. A discursive psychology approach was used to analyze the discourse of flight examiners, based on 7 extended interviews about performance aspects. This study shows that in the discourse of flight examiners, cockpit performance is presented holistically, even though it manifests itself in different ways. Six main discourse repertoires are identified in examiners' discourse about flight deck performance, each of which has between 3 and 5 identifiable sub-dimensions. Case studies show the connectedness and inter-determination of the 6 main repertoires for talking about what pilots do and how they do it.

Key words: performance assessment, non-technical skills, technical skills, linear models, non-linear models.

Assessment is central to maintaining high levels of pilot performance and thereby maintaining safety in the airline industry. The legitimacy underlying performance assessment has seen considerable research and debate into areas such as the conditions under which individuals or teams are assessed (e.g., Bodle, Kaufmann, Bisson, Nathanson, & Binney, 2008; Dawson, 2011),

*Corresponding author. Tel.: +61 7 3735 5604.

E-mail address: t.mavin@griffith.edu.au (T.J. Mavin)

the instrument used during the assessment process (e.g., O'Connor, Hörmann, Flin, Lodge, & Goeters, 2002; Schijven & Bemelman, 2011), how to train those tasked with assessing performance (e.g., Baker & Dismukes, 1999; MacDonald & Sulsky, 2009; Woehr & Huffcutt, 1994), and how assessors make judgments about performance (e.g., Brannick & Brannick, 1989). However, there is little research on the language (discourse) pilots use to describe and classify performance while assessing it. For instance, in the following excerpt, a flight examiner with over 28 years of performance assessment experience is talking to another flight examiner describing a pilot's level of performance during a check.

(i) He did nearly everything well. (ii) He was possibly a little weak on the manipulation side . . . manipulation wasn't one of [his] strong points, but, it was adequate obviously, but it wasn't a starring performance. (iii) But we took off from [departure airport] and just after we got airborne we were notified by ATC that the [destination airport] had a hailstorm and we had hail all over the runway . . . this was on a clearance to line check, it could only happen on that. (iv) He did all the right things, he said "we won't go too far" and held over [navigation aid between departure and destination airports] he made all the right decisions, (v) got all the necessary information, (vi) decided there was no point going on, and return to [departure airport] and did it in a very nice fashion (yea). (vii) Well-managed, (viii) well communicated, kept everyone informed and I was in no doubt what he was thinking in terms of fuel strategies and destination strategies. (ix) We had enough gas on-board, so we had more than one option, so we contacted the company and said which option do you want us to take, where did they want us to go. Yea, did a very nice job. (Examiner 7: 20)

In this example, the flight examiner outlined that he was happy with the pilot's overall performance (i). Whereas the pilot's *ability to fly the aircraft* was described as not exceptional, it was adequate (ii). The examiner talks about the assessment flight—a real flight—where soon after departure air traffic control (ATC) had relayed a message that due to a recent hailstorm, their destination airport was unusable (iii). The pilot under assessment is said to have made a decision to hold mid way between the departure and destination airport (iv) so that more information could be obtained (v), where he finally made a *decision* to return to the departure airport (vi). The description presents the flight as *well managed* (vii) and *communication* between crewmembers was *good* (viii). The examiner, who at that time was acting as first officer, was in no doubt about

the *decision* relating to fuel and destination strategies. This brief analysis shows different dimensions of performance—ability to fly aircraft, communication, and management—and the quality thereof (“*well-managed*,” “*good*,” “*nice job*,” “*nice fashion*”). That is, in the field, the question about the quality of performance appears to be more an issue of classifying performance according to descriptions rather than about applying a measurement instrument and its accuracy. What is the extent of this assessment-related discourse? Which categories do people in the field of aviation use to make sense of and assess performance?

This study was designed to examine the assessment-related discourse of experienced airline pilot examiners. That is, this study is concerned with the everyday practical discourse used to identify aspects of performance and constitute reasons for determining different performances to be of different levels. This discourse is a fundamental resource to justify a pass or fail decision. Grounded in *discursive psychology*, and employing a *grounded theory* method, we identified the interpretive repertoires flight examiners drew upon for talking about the types and quality of performances of pilots undergoing promotion to airline captain.

BACKGROUND

Assessment in Aviation

Historically, pilot training focused on flying skills and associated aircraft knowledge. Even with the introduction of jet aircraft in the 1950s, this training approach has seen little change (Mavin & Murray, 2010). These skills and forms of knowledge, often referred to as *technical skills* (Flin, O’Connor, & Crichton, 2009), have been identified as important to the improvements seen in aviation safety (Johnston, Rushby, & Maclean, 2000). However, the significant decline in aviation accidents over the last 50 years has now changed the percentage mix of pilot- versus machine-related accidents (Helmreich & Foushee, 1993). With improvements in aircraft engineering, design, and technology, pilots have been implicated more frequently in aviation accidents (Flin, O’Connor & Crichton, 2009; Helmreich, 2000; Nagel, 1988). Research into accidents over a period similar to the introduction of the jet aircraft (1959 to 1989) identified pilot error as causal in 70% of worldwide accidents (Helmreich & Foushee, 1993). However, research showed that the majority of pilot errors were not related to technical skills, but, rather, were centered on areas such as communication and decision-making (Helmreich, Merritt, & Wilhelm, 1999). Investigations

revealed a mismatch between traditional training and assessment methods, and causes of accidents, highlighting an important need for changes in pilot training and assessment (Salas, Burke, & Stagl, 2004). Some of these broader methods in pilot training developed under the banner of *cockpit resource management*. Its prime objective was to increase the “timely and proficient use of aircraft resources by operating crew” (Johnston, 1993, p. 371). The training of cockpit resource management, later referred to as *non-technical skills* (NTS), was embraced by airlines and remains foundational in many training programs today. Despite new training programs, the aviation industry found the assessment of NTS difficult. In 1996, the *Joint Aviation Authorities* in Europe recommended the study of NTS assessment, which led to the formulation of detailed NTS assessment measures called NOTECHS. The NOTECHS approach divided specific NTS areas, such as cooperation, management, leadership, situational awareness, and decision-making, into smaller, measurable components. The measures developed within NOTECHS mirrored behavioral descriptors, which presented the assessor with written criteria that matched the performance of the pilot (Flin et al., 2003).

The advance of NOTECHS and other airline-specific assessment approaches was a significant step that improved the totality of criteria required to operate in a complex environment like aviation. Yet the levels of automation in modern aircraft and the complex relationship between technical proficiency and non-technical skills distort the traditional divide that has arisen in aviation (Dekker & Orasanu, 1999; Dekker & Woods, 1999). Robust assessment of pilots’ performance has to emphasize crew resilience (Dekker & Lundström, 2007; Dismukes, Berman, & Loukopoulos, 2007), which in fact requires a better understanding of the assessment process itself. How do senior airline pilot examiners actually assess the performance in the field? What are the features of their discourse used to categorize performance and to qualify differences in performance levels? Such a focus on the assessment language used in the field requires appropriate theory and method, which exists in the discursive psychological approach.

Discursive Psychological Approach

Discursive psychology emerged about 25 years ago with the publication of *Discourse and Social Psychology* (Potter & Wetherell, 1987) and *Discursive Psychology* (Edwards & Potter, 1992). In its turn to the analysis of how people in everyday walks of life use language to conduct their business, it is a distinct way of doing social psychology. Thus, for example, rather than attempting

to measure psychological and cognitive (mental) states of people, it focuses on, for example, how people mobilize psychological concepts as resources for understanding the situations in which they are working and others with which they are dealing. This interest in how people use common-sense psychological concepts has not been attended to in traditional psychology, if it were not for the interest of some scholars in folk psychology (Keil, 2010).

Although some suggest that in its focus on discourse using qualitative methods of investigation, discursive psychology is not a rigorous psychological science, there is evidence across many studies that the projects conducted from this perspective “are scientific: rational, objective, theoretically coherent, empirically-based, and open to public scrutiny” (Stokoe, Hepburn, & Antaki, 2012, p. 487). Discourse psychological studies, though qualitative, may be large-scale, identifying patterns in discourse practices and social interaction that hold across many situations and large datasets.

An important aspect of this approach—theory and method—is the identification of interpretive or discursive repertoires (e.g., Gilbert & Mulkay, 1984; Potter & Wetherell, 1987). A repertoire is a particular form of discourse on which people can draw to support the more or less tenuous claims that are being made in talk. That is, repertoires are not specific to individuals; rather, they constitute inherently shared ways of describing and reasoning about some area of interest. Recent research—in a wide range of fields, including psychology of sports and exercise (e.g., McGannon & Schinke, 2013), marketing (Hall, Gough, & Seymour-Smith, 2013), health (e.g., Jeffries & Grogan, 2012), public understanding of science (Zeyer & Roth, 2013), or community and applied social psychology (e.g., Wilkins, 2011)—shows that individuals, in all walks of life, constitute and support the factual nature of statements, beliefs, attitudes, and epistemologies by means of such interpretive repertoires. It turns out that this process is more revealing than the particular belief, attitude, or epistemology expressed and supported. To exemplify the approach and the core concept of discursive/interpretive repertoire, we use the following situation from one of our databases.

In the following exchange two flight examiners talk about what they have seen following a missed approach procedure during the 1,000 feet prior to reaching missed approach altitude. According to standard operating procedures, there should not be talk that is non-operational at this stage of the flight. The two pilots under assessment are talking about the limited fuel resources they have.

- 1 R: Yeah, so they're talking at an inappropriate time about things they could have left another minute or two.
- 2 L: Correct.
- 3 R: So, management here is not good.
- 4 L: And also, not only Alt STAR capture wasn't mentioned, wasn't identified, you've got a message on the ADU on the top screen saying there's an issue with the autopilot, we have to trim. Something on the autopilot, so it's got to be fixed. It's pretty important because if it's in icing conditions it could manifest itself in a different way. So as you say, it's, you're talking about the fuel, which is important, but at the wrong time.

In this situation, the performance of the flight deck is talked about in terms of communication, which, in this situation, is untimely. As our research reported below shows, communication is one of the discursive resources that flight examiners use to talk about what is happening on the flight deck. The timeliness of communication is one of the modes in which this repertoire expresses itself. Flight examiner R, however, does not just mobilize the communication repertoire but also the management one. Thus, poor management of the order of tasks has led the two pilots to speak at an inappropriate time. Flight examiner L (turn 4) explicitly articulates and summarizes the particulars of this case, drawing again on the communication repertoire and its mode of timeliness. The other parts of turn 4 concern issues that need to be addressed, but, because they cannot be done simultaneously, they have to be organized. Managing the workload would be one of the dimensions of the management repertoire identified in the study below.

A discursive psychological approach is not interested in whether communication and management are real, whether it can be measured accurately, or whether other flight examiners might come to the same rating of these dimensions—i.e., whether there is high interrater reliability—but in the situational mobilization of specific repertoires. Given the solid establishment of discursive psychology as a rigorous science (e.g., Billig, 2012; Edwards, 2012; Potter, 2012), we conducted this study based on the assumption that beliefs, attitudes, concepts, truth, Self, ontologies, sociologies, and epistemologies are integral to socially shared discourses rather than properties of individual flight examiners. In this study, our specific research question pertains to the discourse that flight examiners use to talk about the performance of pilots. To identify the discursive repertoires and the components thereof, we use a grounded theory approach.

METHODS

This study was designed to construct a grounded theory model of the discourse used by flight examiners for talking about the performance of pilots. To identify the discursive repertoires mobilized, we employed a grounded theory approach (Corbin & Strauss, 2008). To us it is a useful approach, precisely because of its orientation toward scientific rigor that is sometimes decried to be “objectivist,” “scientific,” and “positivist” (e.g., Bryant, 2003). The goal of the analysis was to derive a model that completely described the features of the assessment-related discourse within the dataset.

Participants

In this study, seven flight examiners—in some countries referred to as check captains or type-rated-examiners—from a country in the southern hemisphere were interviewed. They aged from 59 to 67 years with a mean of 62.1 ($SD = 2.3$) with a mean of 18,970 hours of total aeronautical experience ranging from 14,000 to 25,500 hours ($SD = 3,700$). On average, they had 24.9 years of experience ($SD = 4.9$) in the role of flight examiner, ranging from 17 to 29 years.

Data Collection

To sample the assessment discourse of flight examiners, semi-structured interviews were conducted lasting between 55 and 120 minutes for a total of 522 minutes. Flight examiners were asked to discuss the assessment of one or more candidates considered for promotion to captain (i.e., a “command upgrade”). The discussed topics included final simulator assessment and the final assessment on the aircraft prior to promotion to captain. The participants were instructed that of particular interest were those difficult situations where the applicant’s performance had turned out to fall near the boundary between a pass and a fail. In all cases, the flight examiners talked about at least three scenarios where the pilot in question had been at the minimum standard level. We anticipated this to elicit justificatory discourse, as our interest was not in *what* assessment they arrived at (pass or fail) but in the reasons flight examiners cited for justifying one or the other assessment. Flight examiners were asked to talk about assessments conducted both on the

simulator and during a routine flight. All interviews were recorded using a digital recorder. The recordings were subsequently transcribed verbatim and checked by a second reader to ascertain the accuracy of the transcription. The total discourse sample consisted of over 99,000 words of assessment-related discourse.

Data Analysis

We began our analysis of the entire database by engaging in cycles constituted by independent analysis of part of the data, followed by meetings in which we compared and negotiated emerging constructions, and discussed new directions for the analysis. This constitutes a “stepwise replication” approach that allowed us to establish the reliability of the findings (e.g., Lincoln & Guba, 1985). We had decided to identify flight examiners’ discursive resources (interpretive repertoires) that they employed in support of their claims about the performance of pilots during an assessment session. We began subsequent meetings by discussing our categories and the examples from the transcriptions that exemplified them. In other instances, one researcher described some phenomenon such as “teamwork” or “leadership” and then the second researcher checked in the entire database to find out all the discursive repertoires employed to talk about the phenomenon. The results were submitted for a crosschecking procedure by the first researcher. Any differences were discussed until resolved. For example, one of us drew a model in which generic arrows pointed up and down; the other proposed a model of the type used here. The first model was more like diagrams that might be observed in the business community, lacking the causal and mediating relationships that a good psychological model should contain. But it had captured the observed hierarchical relations between the different discursive repertoires. From there we worked iteratively, retaining the hierarchical nature but checking the data for connections participants made. We ultimately settled on the approach presented here when we found that during debriefing meetings of a subsequent study, flight examiners gestured over the first model in the very ways that the arrows in our second model describe.

FINDINGS

This study was designed to investigate the assessment-related discourse that flight examiners employ for talking about pilot performance. Six main interpretive repertoires were identified:

situational awareness, decision-making, flying skills, understanding of standard operation procedures, management and communication. Each of these interpretive repertoires are described and exemplified. We then present a unified model of these repertoires grounded in the data.

Six Discursive Repertoires for Talking about the Performance of Pilots

Situational awareness. Situational awareness was perhaps the most fundamental aspect of the assessment discourse. All examiners used it for talking about pilots undergoing promotion to captain. For a pilot who desired to be promoted to captain, it was essential to maintain situational awareness themselves, whilst also sharing their awareness with other crewmembers. Every aspiring captain had to exhibit behavior so that he could be described as being situationally aware. This is apparent in the following quotations.

I don't think I've ever seen anyone who was *really situationally* aware and failed.
(Examiner 2: 166)

If he is not really aware that there is high terrain there, and, there is three or four airplanes coming into this area, the fact that he might be able to fly *doesn't mean diddly-squat* if he's going to run into something. (Examiner 3: 270)

In the second quotation, the examiner talks about “the manipulation side,” the technical performance area on which assessment traditionally has focused. The examiner highlights the fact that the best technical skills will not make up for lack of situational awareness.

For examiners, there existed different components to situational awareness. That is, talk about situational awareness could be differentiated into talk about *perception, comprehension, and projection*. The perception talk related to situations where a pilot was required to obtain information from a variety of sources, both within (e.g., instrumentation) and outside (i.e., environment) the flight deck. Failure to perceive could be due to workload; such as in the following case when a pilot was busy dealing with an engine failure during the landing phase and had forgotten to perceive that the landing gear was still in the retracted position prior to landing.

And I said, “Go around.” So he poured the power all the way and he said, “*Gear up.*” And

I said, “*The gear is up.*” And he nearly died and of course he forgot to put the gear down.
(Examiner 3: 224)

In this case, the examiner had required the pilot to initiate a missed approach (aborted landing) by calling “Go around.” It was when the examiner had said, “The gear *is* up” that the other pilot had finally *perceived* the gear was not correctly positioned for landing. Failure to perceive relevant information frequently was attributed to workload. In another example of the perception sub-repertoire, a pilot was delegating the primary flying tasks to the first officer. Whereas the examiner did not appear to have an issue with delegating primary flying tasks, he used the perception sub-repertoire to talk about the fact that the pilot was not regularly updating himself on primary information.

As soon as there was a problem, he [the pilot in question] would get the other guy [first officer], “*your aeroplane,*” [delegating the task of flying] and he would get the QRH [emergency checklist] out and he’d do all the reading and then [he would] look up and say, “*Oh, where are we now, what is going on?*” Now the captain should never be out of the loop and the captain can’t be out of the loop. (Examiner 3: 257–259)

In this case, the examiner describes the pilot reading the checklist and heard the statement “Where are we now?” as an indication that the pilot was at times not aware of the unfolding dynamics of the aircraft’s flight path.

Comprehension is another part of the situation awareness repertoire. Examiners used it to talk about the inability to correctly comprehend what has been perceived or a fundamental lack of understanding of a system. For example, in the following instance, the pilot is described as navigating the aircraft using a less accurate internal navigation system known as the flight management computer (FMC) and not using the more accurate *navaid* (ground-based navigation system). In so doing, the pilot “missed,” that is, failed to understand, an important flight-related aspect:

He was just relying on the FMC of the aircraft to fly, he was relying on that completely and he’d missed the basic tenet that you have with supporting your [navigation] data up from the navaid itself as a support to confirm whether or not you are actually on the track that

you're supposed to be on. (Examiner 1: 142)

The third component of the situational awareness repertoire related to the projection of future outcomes. In the following excerpt, the pilot had departed from an airport without a full awareness of how another aircraft could conflict with his aircraft's future track.

There was traffic coming inbound and I don't think that he handled that traffic too well. He didn't really separate in a manner that told me that he fully understood where he was and what he was doing. (Examiner 1: 63)

As the aircraft passed each other a little closer than the examiner 1 had hoped, he asked the pilot to look out the window saying, "Just have a look out that you can tell it's [a large commercial jet] going right past us. I think that might have been a little too close!" (Examiner 1: 72). That is, the pilot had not projected into the future that their plane would be close to the large commercial jet.

This excerpt is interesting because it also shows how two or three dimensions of the situational awareness repertoire might be mobilized for describing the same situation in a more differentiated way. Thus, in this quotation the pilot had not projected what would be in the future, and this was derived from the fact that the pilot had not "fully *understood* where he was and what he was doing." In this assessment, the use of the situational repertoire was central to the discussion about pilot's performance.

Decision-making. A second important discursive repertoire consisted in decision-making. It was used to describe situations where a pilot had to resolve specific issues that arose during the flight. In the following example, the examiner talks about a pilot who apparently was reticent to make a decision because he asked a company engineer over the company radio what the implications were for landing the aircraft above its maximum landing weight.

You could say a conservative thing to do. That's nice! But in terms of, he was really just delegating his responsibility for the safety of that aeroplane and all its occupants to somebody who may or may not be qualified to make that decision. (Examiner 2: 34)

Whereas the examiner acknowledged that obtaining further information from various sources was

“a conservative thing to do,” he was concerned that the pilot in question was rather attempting to have the engineer take responsibility for landing the aircraft overweight. In the decision-making repertoire, a captain must acknowledge and accept responsibility for that final decision.

The decision-making repertoire was used to talk about the reluctance of taking responsibility for a decision or for expressing concerns when captains-to-be were either reticent to make a decision or, once a decision had been made, about the confidence associated with it. This did not prevent a subsequent review of the decision; rather, it was more about the confidence shown with the initial decision. In the following excerpt, the examiner draws on the repertoire to talk about a pilot who has to make a choice and then implement a plan of action that would enable the choice to be implemented.

He was tending to be a little bit uncertain of some of his decisions, but what he needed was that it was a good decision or whatever, rather than taking command of it and saying, “Right, grab it by the throat,” and whether I like or not this is what we're going to do.
(Examiner 3:24)

An important dimension of the decision-making repertoire consisted in the sub-repertoires of *fact collection* and *diagnosis*. The discourse was about the ability to obtain specific facts inside or outside the aircraft (“As a captain, you need to take more information on board,” Examiner 6: 8), through a variety of sources, including observation—e.g., looking at instruments or listening to radio calls—or through other sources (e.g., other crewmembers). The following example shows how a pilot made incorrect navigational decisions by assuming that the specific navigational aids were co-located. The pilot was described as having failed to recognize that the two navigational aids were located in different parts of the airport (one navigational aid, VOR, was 1 nautical mile north of the airport whereas the other, NDB, was in the middle of the aerodrome). The implications were the aircraft was correctly positioned, though the pilot had not realized this during the approach to land.

So he didn't work out that the VOR and the NDB are displaced from each other so there was no thought of entering a holding pattern, while he slowed things up. (Examiner 2: 120)

In this situation, the examiner actually mobilizes one of the recurrent dimensions of the decision-making repertoire: the time available for making a decision. In this type of discourse, the pilot could have given himself more time—by entering into a holding pattern—thus allowing for greater time (“while he slowed things up”) to analyse the facts.

Maintaining aircraft within flight tolerances. Flight tolerances constitute a third repertoire. At all times a pilot had to perform such that he could be described as exhibiting good hand manipulation of the aircraft and as maintaining the aircraft within the tolerances that were laid down by the regulatory authority. In the following example, the repertoire is used to talk about a pilot with difficulties conceptualizing the difference between flying a propeller-driven and a jet aircraft.

The operating skills of a jet aeroplane, as opposed to a power-driven aeroplane, the application of thrust in a propeller-driven aeroplane is instantaneous, in the case of a jet aeroplane is obviously less responsive [is slower]. It's about attitude flying, if you can understand where I'm coming from? So the guy would develop high sink rates and couldn't recognize the sink rates; he would think that putting on power would assist, but of course all that did was exacerbate it. (Examiner 6: 20)

Maintaining the aircraft within tolerances could also be achieved with *use of the aircraft's autopilot system*, one of the dimensions of the flight tolerance repertoire. In the following example, a pilot undergoing a command assessment was said to be unable to recognize the problems associated with a degraded automatic throttle system. In this situation, the pilot insisted on using a highly automated descent mode, without recognizing the implications of this.

But the aircraft had a fault, in as much as the auto-throttle was not working properly. The [pilot] had been taught how to do a VNAV descent, and he was going to do VNAV descent with auto throttle disengaged, and from your experience as you know, sort of, the difficulties that brings with the descent profile. He insisted on using the VNAV path descent all the time, and that combined with the fact, that he obviously didn't know about the other parts [autopilot modes] of the descent available and didn't know how to manage them, and didn't know what the lack of auto throttle meant, and for me wasn't up to

command standard. (Examiner 7: 25)

Aviation knowledge. For examiners, the aviation knowledge repertoire consisted of a combination of both technical facts and standard operating procedures. Technical facts could be operational requirements, such as the speed to enter a holding pattern, aircraft limiting weights, or company fuel policy. Standard operating procedures were information taken from company manuals facilitating the orderly process of operating company aircraft, including takeoff techniques, standards phraseology, and sequence to set up aircraft for departure. In the following example, Examiner 1 outlines a pilot that is unaware of the need to enter a holding pattern at 210 knots. “But he just wasn’t aware of it. He just sailed into this thing at 300 knots” (Examiner 1: 217). Prior to this, the same examiner had outlined his concern for the pilot’s lack of technical knowledge by describing the pilot’s misunderstanding of the aircraft flight management computer.

The moment you put a change of runway in the flight management computer it dumps all the information pertinent to the original runway. But [the pilot] totally missed the point that he had the wrong power setting in the computer. It defaults to full thrust setting, so he was erring on the right side, but the fact is he didn’t pick up on how the bloody thing works; I started getting worried straight away. (Examiner 1: 189-190)

This repertoire is in play to articulate the expectation that all pilots operate the aircraft with consistent adherence to standard operating procedures. In one example by an examiner, the pilot undergoing a final assessment was said to have mixed up the engine start procedure in attempting to start an engine without clearance from the ground engineer:

At the end of the check in my mind came up when we pushed back at [Capital city] and he started the engines without a clearance from the ground engineer, and you know, I felt as though it was a dangerous practice, ((laughs)) not a good place to be in. Anyway so yep, in the debriefing of course I had to explain that your not supposed to do that. (Examiner 4: 51)

In this instance, the performance was classified and described in terms of aircraft knowledge generally and the standard operating procedures specifically.

Management. The fifth repertoire identified was used to describe performance in terms of managing operations. Management repertoire includes the identification of tasks, task prioritization, and task delegation: “the importance of things, and prioritizing [them] time wise” (Examiner 4: 14). In the following example, this repertoire is used to talk about a pilot who had noted that one of the aircraft engine instruments was malfunctioning as the aircraft was being taxied towards the terminal. Whereas the examiner conceded that this was important, the management repertoire allowed justifying the recommendation that the first officer taxi was not the most appropriate way to delegate tasks.

We were taxiing in—this once again was a two-day event—we were taxiing in and the idle RPM on one engine was a bit higher than the other, the N1 that is. And he actually got a book out and looked up what it should be and you know, it was a bit high, he was quite right, it was but I thought with flight management wise, when you’re taxiing in, it’s not a good time to be getting books out and checking the N1. . . . You know, he could have just written down what it was or made a mental note of what the actual reading was and looked it up in five minutes when he put the park brake on, that would have been more sensible. (Examiner 4: 171-173)

Communication. This repertoire was employed to describe those situations that required the sharing of information. For example, an examiner might say: “This is communication, this guy is not communicating. He didn’t say anything to me whatsoever” (Examiner 1: 223). In such a situation, it is the lack of communication that was used to characterize pilot performance. The examiners recognized that communication is difficult to assess and to teach.

The harder things to define are the things like the you know, the level of situation awareness, the flight management, the ability to communicate well with your co-pilot and with the rest of the team in a sense of air traffic control and engineers and particularly the cabin supervisor. So we play that scenario in the simulator and you’re really looking for the skills of the person under assessment in that area, um and coping with um those sort of relationships under pressure, under stress if you like. (Examiner 2: 13)

As a result, examiners tend to play scenarios that not only test communicative performance but also teach communication especially in situation when the workload is high and the pilots are under stress. In some instances, the repertoire concerned the clarity of the information transfer between crewmembers. For example, in the following excerpt, it is used to discuss a young pilot who was not clear in his communication (“I can’t hear,” “he mumbles,” “am I deaf”).

My criticisms were the same every time: “I can’t hear what you’re saying, I’m getting old and I’m a bit deaf maybe.” But he tends to mumble to himself, and he [isn’t] forthcoming with a lot of discussion about developing circumstances. (Examiner 1: 376)

Other aspects of the communication repertoire concerned the timeliness of the information sharing. In the following example, the repertoire is used for talking about the performance of a pilot who had decided to divert to another airport during a simulator exercise. However, he had failed to communicate to the first officer (“there was no reference to”) details of why he was going to divert.

In that case, *there was no reference* to how much fuel he had left. He probably knew how much fuel he had left but he didn’t mention it to the first officer. [What he should have said was] I have got “x” amount of fuel left and we are going to burn this to get to where we are going, so therefore we are okay, and that was never mentioned. (Examiner 3: 36)

A Model of Airline Pilot Performance (MAPP)

During the descriptions of performance, the examiners rarely talked about each in isolation, but rather invoked two or more repertoires. In fact, in the entire database of assessment discourse, performance was never considered in terms of isolated skills that could be measured independently. Thus, underlying the talk is a holistic view of pilot performance that is described in terms of the six fundamental repertoires and the level of quality (e.g., failure to exhibit performance component, low quality). There was always talk about multiple components of performance (our opening example), although situational awareness was the most important dimension of all. We express this in a grounded theory model of assessing pilot performance (MAPP) (Fig. 1a). In this model, situational awareness takes an integrative role, because, as the

examiner above suggests, it cannot be compensated by excellent flying skills. Accordingly, it has been placed at the top of the theoretical model. To guarantee the safety of the airplane, flying it within given tolerances is of greater importance in examiners' discourse than communication, for example. The lines that connect the different repertoires express the fact that any one component of performance could be said to have been affected by another performance component. In the following, we provide an example of how MAPP can be used to model the assessment talk about pilot performance.

«««««« Insert Fig. 1 about here »»»»»»

In the example, an examiner narrates a critical incident from his experience. In this excerpt, the examiner is describing a simulator assessment where the pilot in question is having difficulty understanding the prompts that are being relayed to him by the cabin crew [here described as the "girls"] that they were unable to put a fire out in the galley of the aircraft. The narrative describes performance by drawing on different repertoires of the integrated model of pilot performance (Fig. 1b).

And another one, ah a guy had a fire in the rear galley, and the girls were giving him prompts, "We just fired the forth bottle into this thing and it was still going." That should be a real flag to the guy to get this thing on the ground pronto because we have got an uncontrollable fire. But he allowed air traffic control to push him all around the place; and he was even going a lot slower than he should have been anyway before the approach, and with no urgency, and no forward planning that "we are going to stop on the runway" or are "we going to have the firies [airport fire engines] standing by on a certain point on the runway" are we going to "stop right beside them so that they can get on board and get this thing out." (Examiner 3: 49)

The following components are salient in this narrative.

- Situational awareness (perception) and decision-making (facts): And another one ah a guy had a fire in the rear galley, and the girls were giving him ah prompts, "We just fired the forth bottle into this thing and it was still going."

- Situational awareness (comprehension) and decision-making (diagnosis): That should be a real flag to the guy to get this thing on the ground pronto because we have got an uncontrollable fire.
- Situational awareness (projection) and decision-making (risk): to get this thing on the ground pronto because we have got an uncontrollable fire
- Management (task and task prioritization): But he allowed air traffic control to push him all around the place; and he was even going a lot slower than he should have been anyway; with no urgency and no forward planning

This situation can be expressed in terms of the model in the following manner. The poor situational awareness affected decision-making and management. Moreover, as the pilot did not appear to appreciate the fact that the fire constituted a considerable risk, he did not diagnose the situation sufficiently, and, as a result, did not assert himself with respect to the control tower and was flying too slow. That is, poor situational awareness and decision-making affected management. The relations between the different performance components are shown in Fig. 1b. The two repertoires not mobilized in the narrative are communication and aviation knowledge.

Member Checking

The member checking procedure is used to ascertain that a grounded theory model derived through the analysis of the data reflects the participants' reality and experience (Lincoln & Guba, 1985). In this study, the MAPP (in its original first model state) was presented to flight examiners during a conference presentation. That the model represents the reality of and is pertinent to the field is acknowledged in the fact that as a result of the presentation, three airlines of the southern hemisphere implemented it as the focal point of its crew resource management training and as the core of the assessment process. Furthermore, a major military organization has now adopted the MAPP as the main framework for technical and non-technical skills training. That is, the chief examiners, training managers and military personal recognized that the MAPP represented their everyday reality of assessing pilots. As our ongoing ethnographic work shows, flight examiners consistently use the MAPP to talk about areas of weakness and strength in the performance of pilots. Moreover, they articulate, for example, how the performance qualities in one or two areas mitigate the performance quality in another. Examiners tend to point to the relevant areas so that

their hand/finger gestures trace out figures similar to the way arrows are used in Fig. 1b. Thus, in a recent debriefing session that was observed, the pilot being assessed had begun to turn onto a 15-mile arc rather than the required 10-mile arc during his approach (he was going to turn too early). The flight examiner talked about the incident in terms of lowered situational awareness, which was mitigated by the pilot verbalizing the anticipated turn (communication) to the other pilot who was able to correct this misunderstanding. From here, he was then able to organize his task sequence to exit the situation (management). While providing this description, the flight examiner pointed to situational awareness, then to communication, back to situational awareness and finally to management (Fig. 1c). He then explained that it was “good communication” and “excellent management” that had allowed the pilot to recuperate his situational awareness. The final outcome was that the aircraft could be maintained within flight tolerances.

DISCUSSION

This study was designed to investigate the assessment-related discourse of flight examiners. Six main interpretive repertoires were identified, each consisting of more fine-grained sub-repertoires. Although examiners talk about different aspects of performance, their discourse treats it in a holistic way. Thus, performance almost never is talked about using a single repertoire. Rather, other repertoires are often used to explain, for example, why situational awareness was lowered—e.g., the task was not managed well, leading to an increased workload, which affected situational awareness. A grounded theory model was developed and represented in diagrammatic form (Fig. 1). The discursive psychological approach employed to frame the study is not concerned with the match between assessor categories and true pilot performance. Rather, it is concerned with understanding and classifying the discursive resources upon which examiners draw for supporting their pass/fail decisions in their daily practice.

The results of this study show that the flight examiners’ discourse treats performance holistically without separating it into technical and non-technical aspects. The analysis of flight examiners’ discourse therefore suggests that rather than separating performance into technical and non-technical skills (e.g., NOTECHS), an integrated approach, as in the model presented (Fig. 1), better reflects the reality of everyday pilot assessment as we find it in practice. However, although all performance aspects are important, examiners give greater weight to the performance dimensions of situational awareness, aircraft flown within tolerances, and decision-making—

consistent with the literature on non-compensatory skills (e.g., Brannick & Brannick, 1989)—which is reflected in their greater reliance on the associated discursive repertoires. In the following, we discuss the findings under two aspects: (a) the reality of practical performances and its reflection in measurement and categories of understanding and (b) the explanation of classical performance concepts absent in the present model.

Real Performance, Measurement, Categorization

The performance assessment literature often raises questions about reliability and provides measures for the degree to which two or more assessors agree (e.g., Bodle et al., 2008). Associated with the discussion of reliability is the notion of rater subjectivity. Discursive psychology approaches the issue very differently and, in so doing, sidesteps questions about the reality of such dimensions such as situational awareness (e.g., Dekker, Hummerdal, & Smith, 2010; Van Dijk, Van Der Merwe, & Zon, 2011) or the difficulty that scientists find categorizing such concepts (e.g., Saurin, Wachs & Henriqson, 2013). By asking questions related to the usefulness of particular discourses in dealing with the everyday world. Even if raters differ in the way in which they categorize performance—e.g., whether it was *knowledge* or *situational awareness* that was “weak” in a particular situation—the interpretive repertoires are never at stake. These are taken as the ground that never comes to be questioned in pilot–assessment-related discourse. That is, the discursive approach is concerned with the ways in which everyday people, such as flight examiners, constitute and talk about the worlds of their experience. It is also concerned with the ways in which categorical terms (e.g., “good,” “poor,” “minimum standard”) are used to assign qualities to the performance dimensions. This aspect is further made salient as our ongoing research shows.

The three airlines using the model of pilot performance use evaluation grids that consist of categories of performance (vertical dimension) and the quality thereof (horizontal dimension). The quality is expressed in “word pictures.” Thus, for example, in the repertoire of situational awareness, there are the sub-repertoires of perception, comprehension, and projection, as in the classical model of situational awareness (e.g. Endsley, 1995). For the projection sub-repertoire, there are five “word pictures”: (a) did not predict future events, even those obvious to flight safety; (b) difficulty predicting future events; (c) some difficulty predicting future events; (d) predicted future events and impact on flight safety; and (e) actively considered future events and impact on

flight safety. That is, the quality of a performance is expressed in verbal form (i.e., a discourse). The assessment of the nature and quality of performance, therefore, has become a categorization issue rather than being a measurement issue (Roth & Mavin, 2013, in press). The five word pictures are associated with other, quality-indicating categories: unsatisfactory, minimum standard, satisfactory, good, and very good (excellent). Although some research suggests that an increase in the number of categories should decrease assessment variance and, thereby, increase reliability (e.g., Brannick, Prince, & Salas, 2002), the model presented here would suggest that such a move increases the categorization possibilities and, with them, the ranges of sub-repertoires brought to any one issue. This, therefore, would be consistent with the findings of another study, which suggested that there is “no guarantee that the sum of the parts is equal to the whole” (Bodle et al., 2008, p. 212).

Traditional Assessment Concepts Missing from MAPP

Some readers might point out that the grounded theory model (MAPP) developed in this study does not contain explicit concepts often used, such as *teamwork* or *leadership*. However, the flight examiners emphasize that being a pilot means being a team player; and being a captain means being able to manage the workload of both pilots and communicate well with other crew. This is evident in the following excerpts, where examiners speak about what makes a good captain.

It’s a team game. . . There is no room for one-man players. . . . It is not an easy seat to sit in the right hand seat and [the captain] has gone through that, so therefore he should have more compassion and understanding of what the requirement is from being a leader to be able to bring the rest of the team as part of his team. (Examiner 3: 312–314)

That’s right. It’s more how well, it’s um being, I guess, aware of what’s required, and communicating properly with the FO processes in the organization so that making it happen, and then communicating that effectively with the rest of the team. (Examiner 5: 92)

That is, talk about leadership and teamwork tends to be in terms of a mix from the repertoires. To use an analogy, then, MAPP constitutes something like a collection of primary colors from

which other colors can be produced. Teamwork therefore defines the required color, and consequently, each specific situation—say normal verse emergency—may require a differing blend of primary colors to satisfy the situation. For example, a captain who faces a scenario where his lack of knowledge (knowledge) of a system may be impeding his ability to diagnose the situation (situational awareness and decision-making) may ask the first officer for his/her take on the situation (communication) or delegate (management) the flying of the aircraft (aircraft maintained within tolerances) so a manual (knowledge) can be consulted. This is good teamwork, yes, though described using a selection of the six repertoires. The repertoires described above constitute a set of fundamental discursive means to talk about every aspect of performance in the cockpit.

In a similar way, leadership was talked about in terms of a mixture of the categories that made up the grounded theory model (MAPP). In the following example, communication (“Going to each cabin crew, ‘Are you all right?’”), decision-making (“accepting responsibility for *what’s going on around you*”), and management (“command management”) are the three repertoires that are used to describe leadership:

Go to each of the cabin crew “Are you all right?” and it’s not, you’re not doing that because you’re wearing the bars and things, it’s just because, you know, you assume that role, that these things are my responsibility. So you’re accepting responsibility for that, of what’s going on around you, and that’s, you know, that’s command management. It’s, in my airline terms, it’s leadership isn’t it? (Examiner 2: 272)

In this way, the interpretive repertoires identified do in fact cover the same terrain that other specific concepts do. However, these concepts are the result of drawing on a combination of interpretive repertoires.

A possible limitation of this research might be thought to exist in the small number of participants. The limitations of this study, however, are not assessable in terms of number of participants because this study takes *discourse* as its object. This means that it focuses on the structure of a form of language used in the industry and inherently shared within the culture. Nevertheless, it is always possible that additional subcomponents might be identifiable when the number of interviewees is increased. Since the original submission of this study we conducted two additional ones focusing on assessment (Mavin, Roth, & Dekker, 2013; Roth & Mavin, in press)

and on debriefing (unpublished data). In both studies we asked participants for any additional dimensions of actually observed (rather than remembered) performance (in videos or simulator) that were not covered by the model presented here. Neither flight examiners nor assessed pilots could think of any component that had been left out.

CONCLUSION

In this study an analysis of flight examiners' discourse related to pilot performance assessment is presented. A grounded theory model is presented that highlights the integrated nature of performance. In any particular performance, only some of the interpretive repertoires may be mobilized for its description and for justifying the assessment (score). Informal feedback from the three companies employing the model suggests that it assists them in identifying problem areas and in designing initial and remedial training for pilots with problems. Empirical studies are planned to ascertain the degree to which the model assists examiners and trainers to identify and rectify problematic areas of performance.

REFERENCES

- Baker, D. P., & Dismukes, R. K. (1999). Training pilot instructors to assess CRM: The utility of frame-of-reference (FOR) training. In R. Jensen (Ed.), *Proceedings of the 10th International Symposium on Aviation Psychology* (pp. 291–300). Columbus, OH: Ohio State University.
- Billig, M. (2012). Undisciplined beginnings, academic success, and discursive psychology. *British Journal of Social Psychology, 51*(3), 413–424.
- Bodle, J. F., Kaufmann, S. J., Bisson, D., Nathanson, B., & Binney, D. M. (2008). Value and face validity of objective structured assessment of technical skills (OSATS) for work based assessment of surgical skills in obstetrics and gynecology. *Medical Teacher, 30*, 212–216.
- Brannick, M. T., & Brannick, J. P. (1989). Nonlinear and noncompensatory processes in performance evaluation. *Organizational Behavior and Human Decision Processes, 44*, 97–122.
- Brannick, M. T., Prince, C., & Salas, E. (2002). The reliability of instructor evaluations of crew

- performance: Good news and not so good news. *International Journal of Aviation Psychology*, 12, 241-261.
- Bryant, A. (2003). A constructive/ist response to Glaser. *Forum Qualitative Sozialforschung/ Forum Qualitative Social Research*, 4(1). Accessed June 9, 2013 at <http://nbn-resolving.de/urn:nbn:de:0114-fqs0301155>
- Corbin, J., & Strauss, A. L. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Los Angeles, LA: SAGE.
- Dawson, S. (2011). Perspectives on performance assessment in medical simulation. *The Surgeon*, 9, S21–S22.
- Dekker, S. W. A., Hummerdal, D. H., & Smith, K. (2010). Situation awareness: Some remaining questions. *Theoretical Issues in Ergonomics Science*, 11, 131–135.
- Dekker, S. W. A., & Lundström, J. T. (2007). From threat and error management (TEM) to resilience. *Human Factors and Aerospace Safety*, 6, 261–274.
- Dekker, S. W. A., & Orasanu, J. M. (1999). Automation and situation awareness: Pushing the research frontier. In S. W. A. Dekker & E. Hollnagel (Eds.), *Coping with computers in the cockpit* (pp. 69–85). Aldershot, UK: Ashgate.
- Dekker, S. W. A., & Woods, D. D. (1999). Automation and its impact on human cognition. In S. W. A. Dekker & E. Hollnagel (Eds.), *Coping with computers in the cockpit* (pp. 7–27). Aldershot, UK: Ahsgate.
- Dismukes, K., Berman, B. A., & Loukopoulos, L. D. (2007). *The limits of expertise: Rethinking pilot error and the causes of airline accidents*. Aldershot, UK: Ashgate.
- Edwards, D. (2012). Discursive and scientific psychology. *British Journal of Social Psychology*, 51, 425–435.
- Edwards, D., & Potter, J. (1992). *Discursive psychology*. London: Sage.
- Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, 37, 65–84.
- Flin, R., Martin, L., Goeters, K., Hoermann, J., Amalberti, R., Valot, C., & Nijhuis, H. (2003). Development of the NOTECHS (Non-Technical Skills) system for assessing pilots' CRM skills. *Human Factors and Aerospace Safety*, 3, 95-117.
- Flin, R., O'Connor, P., & Crichton, M. (2009). *Safety at the sharp end*. Aldershot: Ashgate.
- Gilbert, N., & Mulkay, M. (1984). *Opening Pandora's Box: A sociological analysis of scientists' discourse*. Cambridge: Cambridge University Press.

- Goldsmith, T. E., & Johnson, P. J. (2002). Assessing and improving evaluation of aircrew performance. *The International Journal of Aviation Psychology*, *12*, 223–240.
- Hall, M., Gough, B., & Seymour-Smith, S. (2013). Stake management in men's online cosmetics testimonials. *Psychology & Marketing*, *30*, 227–235.
- Helmreich, R. L. (2000). On error management: Lessons from aviation. *British Medical Journal*, *320*, 781–786.
- Helmreich, R. L., & Foushee, H. C. (1993). Why crew resource management? Empirical and theoretical bases of human factors in aviation. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 3–46). San Diego: Academic Press.
- 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.
- Jeffries, M., & Grogan, S. (2012). “Oh, I’m just, you know, a little bit weak because I’m going to the doctor’s”: Young men’s talk of self-referral to primary healthcare services. *Psychology & Health*, *27*, 898–915.
- Johnston, A. N. (1993). CRM: Cross-cultural perspectives. In E. L. Wiener, B. G. Kanki & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 367-398). London: Academic Press.
- Johnston, A. N., Rushby, N., & Maclean, I. (2000). An assistant for crew performance assessment. *The International Journal of Aviation Psychology*, *10*, 99–108.
- Keil, F. C. (2010). The feasibility of folk science. *Cognitive Science*, *34*, 826–862.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- MacDonald, H. A., & Sulsky, L. M. (2009). Rater formats and rater training redux: A context-specific approach for enhancing the effectiveness of performance assessment. *Canadian Journal of Behavioural Science*, *41*, 227–240.
- 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, NL: Springer.
- Mavin, T. J., Roth, W.-M., & Dekker, S. W. A. (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.
- McGannon, K. R., & Schinke, R. J. (2013). “My first choice is to work out at work; then I don't feel bad about my kids”: A discursive psychological analysis of motherhood and physical

- activity participation. *Psychology of Sport and Exercise*, 14, 179–198.
- Nagel, D. C. (1988). Human error in aviation operations. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in aviation* (pp. 263–303). London: Academic Press.
- O'Connor, P., Hörmann, H. J., Flin, R., Lodge, M., & Goeters, K.-M. (2002). Developing a method for evaluating crew resource management skills: A European perspective. *The International Journal of Aviation Psychology*, 12, 263–285.
- Potter, J. (2012). Re-reading Discourse and Social Psychology: Transforming social psychology. *British Journal of Social Psychology*, 51, 436–455.
- Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour*. London: Sage.
- Roth, W.-M., & Mavin, T. J. (2013). Assessment of non-technical skills: From measurement to categorization modeled by fuzzy logic. *Aviation Psychology and Applied Human Factors*, 3, 73–82.
- Roth, W.-M., & Mavin, T. J. (in press). Peer assessment of aviation performance: Inconsistent for good reasons. *Cognitive Science*.
- Saurin, T.A., Wachs, P. & Henriqson, É. (2013). Identification of non-technical skills from the resilience engineering perspective: A case study of an electricity distributor. *Safety Science*, 51, 37–48.
- Salas, E., Burke, C. S., & Stagl, K. C. (2004). Developing teams and team leaders: Strategies and principles. In D. V. Day, S. J. Zaccaro & S. M. Halpin (Eds.), *Leader development for transforming organizations* (pp. 325–355). New Jersey: Lawrence Erlbaum Associates.
- Schijven, M. P., & Bemelman, W. A. (2011). Problems and pitfalls in modern competency-based laparoscopic training. *Surgical Endoscopy*, 25, 2159–2163.
- Stokoe, E., Hepburn, A., & Antaki, C. (2012). Beware the “Loughborough School” of social psychology? Interaction and the politics of intervention. *British Journal of Social Psychology*, 51, 486–496.
- Van Dijk, H., K. Van Der Merwe, & R. Zon. (2011). A coherent impression of pilots’ situation awareness: Studying relevant human factors tools. *The International Journal of Aviation Psychology*, 21, 343–356.
- Wilkins, A. (2011). Community and school choice: Geographies of care and responsibility. *Journal of Community & Applied Social Psychology*, 21, 1–13.
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative

review. *Journal of Occupational and Organizational Psychology*, 67, 189–205.

Zeyer, A., & Roth, W.-M. (2013). Post-ecological discourse in the making. *Public Understanding of Science*, 22, 33–48.

Caption

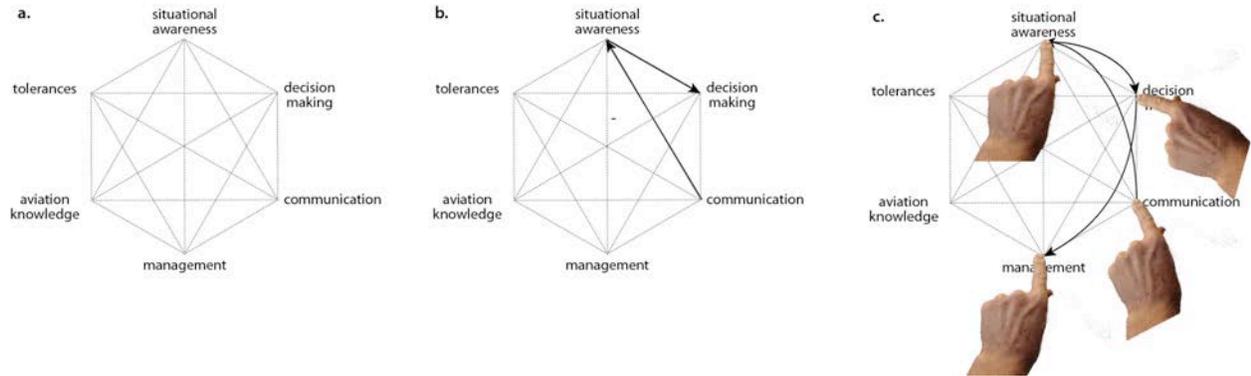


Fig. 1. a. A grounded theory model of flight examiner discourse concerned with the assessment of pilot performance. b. Cabin crew communicates to the captain about an evolving emergency in the form of a fire in the cabin of the aircraft. A pilot does not recognize the danger arising from a fire in the passenger section, which affects decision-making. c. An examiner gestures while explaining how the communication should have altered the captain situational awareness. He gestures how this misunderstanding effected decision making, and, finally, how management of tasks and priorities of these tasks should have been used to recuperate the current circumstances.