# The role of context in the assessment of operational skills: a case in Air Traffic Control

**Kalina Yacef* and Leila Alem****

(*) Thomson Radar Australia Corporation and
Laboratoire d'Intelligence Artificielle de Paris 5

(**) CSIRO Division of Information Technology

Locked bag 17, North Ryde 2113 NSW, Australia
[kalina,leila]@syd.dit.csiro.au

Ph (61-2) 325 3146
Fax (61-2) 325 3101

**Abstract**:

We are interested here in the assessment of operational skills for training purposes. It has been established that this assessment should be performance-based [Alem et al. 96]. We propose to go one step further by analysing these performances in context in order to assess the learning. We describe why the notion of context is necessary in the assessment of operational skills, how it is characterised and modelled, and how the student's performance is analysed in context in order to assess the learning. These considerations are illustrated in the context of the ATEEG system, a simulation-based learning environment for Air Traffic Controllers [Alem & Keeling 96].

## 1. Introduction

The domain of application for the research presented here is the training of Air Traffic Control (ATC) operational skills. These skills, which are acquired after declarative and procedural knowledge, ensure the correct use of these two types of knowledge regardless of the specificities of the situation that the operator is confronted with, and ensure that the expert operator is never snowed-under with unseen situations. ATC work, of which the essential security reasons are obvious, requires not only care and rigour but also a true expertise, fruits of a long experience that cannot be reduced to simple and exhaustive methods. This lack of formalisation gives rise to great difficulty in finding an adequate training method and evaluating the student in the development of his expertise. ATC is a high-performance activity, so indeed performance measures must be part of the student's assessment as it has been established by the research community working in the area of simulation-based learning [Alem et al. 96]. We propose to extend this assessment to a more cognitive evaluation of the learning by using the notion of context in which performance measures were done.

Assessing the learning state is beneficial for instructional purposes: if the tutor knows what the student does or not know, and where he is in his learning process, then the generation of training exercise, of explanation are more adapted. There has been a lot of work on cognitive evaluation/diagnosis of declarative and procedural knowledge, as for example Steamer [Hollan et al. 84], Sophie II [Brown et al. 82], Quest [White & Frederiksen 87], LISP Tutor [Reiser et al. 85], Geometry tutor [Anderson et al 85]. But less attention has been given to the assessment of operational skills: GT-VITA [Chu et al. 95] addresses the training of NASA satellite ground operators but the assessment of their skills is limited to the success of specific procedures. In our system ATEEG [Alem et al 96] we have completed a performance-based assessment of operational skills [Yacef & Alem 96a], and we are currently extending it towards a cognitive assessment of the learning, by analysing these performances in context.

Context allows us to interpret performance measurements because it plays a key role in the development of operational skills: if the trainee always learns similar procedures in similar situations, he/she may grasp the procedural knowledge but if something changes in the task environment, if some part of the reasoning has to be modified, he/she is likely to have difficulty adapting. As experience develops, training objectives aim more and more at facing the learner with increasingly complex and various situations to look at how well he/she copes with them. Hence when assessing such skills, the need is to assess the experience of the learner, his degree of adaptability to new situations and identify his misconceptions. ATC activity is situated [Gaillard 92], which means that any action is a result of the resources and constraints characterising the current context [Suchman 87]. Thus by considering the pertinent elements of this context, we can better understand the student's performances.

In the next section, we go into more details about the training of operational skills. In section 3 we define the criteria needed for assessing them and justify the need for representing the context to support this assessment. Section 4 and 5 describe how the context is modelled and used for analysing the student's performance in ATC.

## 2. ATC operational skill training

Operators of complex and dynamic systems have three main classes of knowledge: declarative knowledge, procedural knowledge, and operational skills [Chu et al 95]. Operational skills are about the ability to apply the two previous types of knowledge in a timely and accurate manner and the ability to cope with any problem-solving situations, including unseen ones [Woods 88].

It is known that certain human decision-making activities can be accurately analysed and represented in the form of algorithms: in particular these concern the tasks of solving "well-defined" problems. In this case it is possible to use the algorithms as a reference for the instruction or for the evaluation of the student. Meanwhile ATC, as many other human activities, depend rather on solving "ill-defined" problems: in this category, although many people exhibit competence and success, their reasoning and the implicit rules that they use can only be insufficiently explained and thus cannot be formalised [Cannat et al. 87; Bisseret et al. 81; Alem & Keeling 96] .

To understand the nature of ATC skills, it is useful to conceptualise the human performance as two levels of knowledge processing [Regian & Schneider 90; Schneider 88]: controlled processing, which requires reasoning, and automated processing which is the result of highly compiled knowledge or procedure and which is put into operation by direct mapping to the context [Boy 91][1]. Controlled processing is slow, effortful, sequential and capacity limited. It is expected at the beginning of practice on any novel or unexpected task. Automated processing is fast, parallel, fairly effortless, not under direct operator's control, and is used in performing well-developed skilled behaviours. These two type of processes interact and are present in any type of complex operations. At a high level for example, an air traffic controller facing an emergency situation or a new conflict would use automated components in handling all the routine tasks, analysing the traffic situation, and controlled components for planning the traffic and handling the emergency or resolving the conflict. His thinking would be focused on the emergency or conflict but all the other routine tasks, including acquiring the information he needs from the

---

[1]The types of knowledge representation underlying these processes are called situational and analytical knowledge by Boy [91], and the corresponding human behaviours are respectively skill-based behaviour and rule+knowledge-based behaviours [Rasmussen 83].

radar screen or communication, are automated tasks. As the experience develops, the number of automated skills increase and therefore the controlled memory is more available to focus on novel problems.

Expertise in operational skills can therefore be seen as a rich and correct set of automatic components as well as a good quality of controlled processing, ie. a set of cognitive resources allowing to deal with new situations and planning. In order to assess this expertise, we need first to define the elements that are crucial to this evaluation.

## 3. Criteria for assessing operation skills

Acquiring these skills is therefore about building automatic components and enhancing the potential of cognitive resources. Both of these broad objectives are reached with intensive practice on training exercises of various type and complexity. The more components are automated, the more processing capacity is freed up and may be applied to controlled components. To attain a high level of automaticity, the student needs to apply his/her knowledge in a large number of situations, and then once reaches a performance criterion, he/she needs to practice under an increasing resource load. Since the ATC tasks are very interwoven, the student naturally practices several tasks at the same time (multi-component training). At the beginning of the training or when learning a new concept, the learner will face adequate situations, at a low workload so that he can start building the framework for automated components from the currently controlled ones. Later, when the component has been automated optimally, he will need to practice under higher workload, in order to fix this automated component and learn to use it in any situation, while performing other tasks (see fig.1). Assessment is needed to know whether the learner is ready to move on or if he needs to go back one step. The breakdown per task is important in order to assess automaticity on a component: the student must be able to execute the other tasks in parallel.

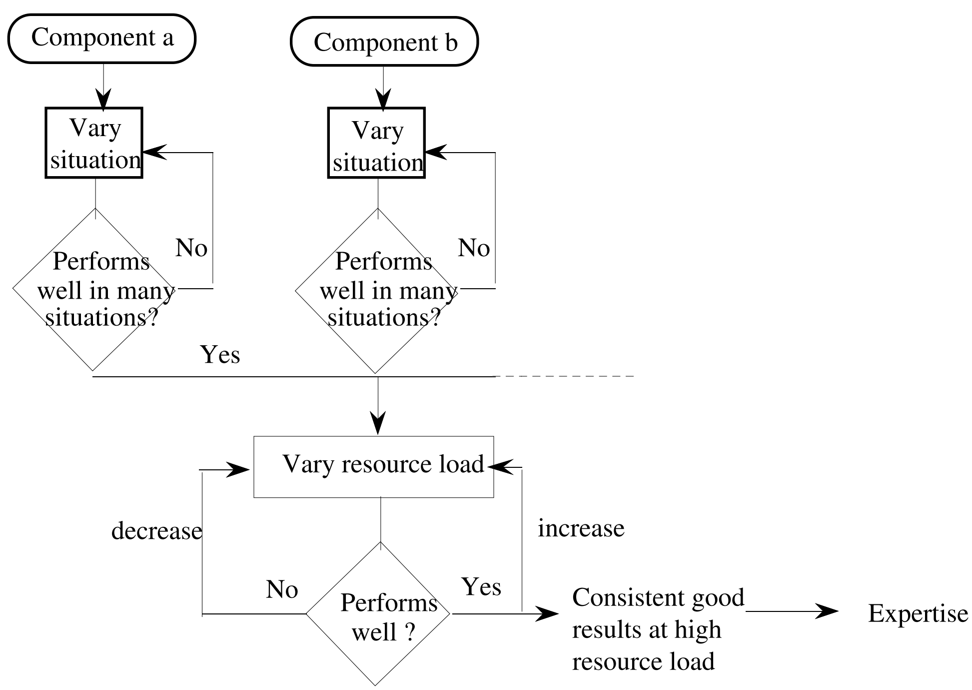


Fig 1. Training process

This process provides a basis for the assessment of the learning and for the generation of instructional strategies. We propose the following steps:
- interpret the learner's performance in context. ATC being a high-performance activity, assessment obviously implies a set of performance measures that can be compared with expert criteria. However performance measures, which generally measure how quickly and/or how accurately a task was done, cannot be interpreted alone. For example, an average performance does not have the same meaning if it was measured on a brand new situation or on a familiar one and the knowledge of the characteristics of this situation may help in pointing out problem areas. Similarly, the workload is important: if the student resolved a conflict safely but not in the most economic way, then the fact that he was under a particularly high workload will indicate a rather good performance (the student was wise enough not to lose too much time on a task). Situation and workload are therefore two major factors.

- the stability of skill regardless of the situation. Once the student is good at a task in any situation, then he/she must learn to do it in less time, ie under higher resource load. It is therefore useful to evaluate whether the student has faced and performed well in a lot of situations. This step also highlights situations that seem to present a problem for the student.

- the level of automaticity. Automated components provide a more reliable performance under stress, high workload, or any circumstances affecting the thinking potential [Schneider 88], as opposed to controlled components which show a drop in performance at high resource load. Observing the level of performance as resource load increases allows us to diagnose the student's level of automaticity.

In ATC training, the elements of the context that are relevant to the assessment are (1) the characteristics of the traffic situation and (2) the resource load, which is an indicator of the availability of the student to perform tasks. This latter is a combination of workload (incurred by the traffic) and mental load (level of stress put on the student).

The next section describes how these concepts are modelled.

## 4. Modelling the Context in ATC

The context is seen here as the pertinent information of the training exercises that can be used for the assessment of the learning. As described above, the context must contain information about the traffic situation and the resource load. We have chosen to focus here on the task of conflict handling (detection and resolution), which is one of the most complicated and representative task of the ATC. Let Ci be the context for a 5-minute interval, Si the traffic situations considered, and Ri the resource load during that interval of time. Interviews with ATC instructors and literature research in ATC training lead us to consider the context as a situation and a certain resource load:

$$Ci = (Si, Ri)$$

The parameters of the *situation* Si provide elements of the traffic that are pertinent to the assessment. In the case of conflict handling between 2 aircraft, the elements of the situation are:
- $\alpha$: the convergence angle between the two aircraft.
- $\varphi$: their respective flight phases, both cruising, both climbing, both descending, climbing and cruising, descending and cruising, climbing and descending.
- $\sigma$: the comparison between their performance/speed: similar performance, quite different performances, very different performance.
So the situation Si is defined as

$$Si=(\alpha, \varphi, \sigma)$$

Example: Two scheduled flights, Boeing 747, in opposite direction, one climbing, one cruising, would generate the situation S = (180°, climb/cruise, similar performance).

The *resource load* Ri put on the student is composed of a workload Wi and a Mental load, Mi. The *workload* Wi is associated to the situation the student is facing. It is determined on a 5-minute interval and reflects :
- the number of aircraft flying at the same time in the sector during the interval
- the average complexity of the aircraft. A controller who is handling 6 "problem" aircraft (ie that are in conflict or emergency, that don't fly a normal route etc..), would have a much higher workload than a controller handling 20 "problem-free" aircraft (ie all cruising, separated, flying their normal route). This complexity is calculated according to the following chart (simplified from measures used at ENAC (French National School of Air Navigation) for an enroute sector)

| Characteristics (static and dynamic) | Workload value |
|---|---|
| Scheduled flight | 1.00 |
| Charter flight | 1.10 |
| Military flight | 1.20 |
| Climbing/descending | +0.24 |
| Coordination | +0.26 |
| Conflict resolution | +1.40 |
| Holding | +0.60 |

Fig 2. Workload table

Example: if we use data from the previous example, the workload generated by these two aircraft would be:

W = (Scheduled flight) + (Scheduled flight+climbing) +(conflict resolution)
= (1) + ( 1+ 0.24 ) +( 1.40 )
= 3.64

And if there are 2 other scheduled cruising aircraft, not in conflict, in the same sector, the total workload would be:

W = 3.64 + 1 + 1 = 5.64

It must also be noted that an increase in the workload not only raises the level of controller's activity but also naturally restraints the number of solutions available to a conflict, making the resolution more difficult.

The *Mental load*, Mi, represents the elements not derived from the traffic itself and which have an impact on the working capacity of the student. The stress from the environment (noise, poor quality of communications etc..) as well as the past of the exercise have an impact on the controller's mental load. At this stage we consider only two elements of the exercise:
- the length of the exercise is an important factor. A student who has been working for 1 hour on an exercise has a greater mental load than the one who has just started. Instructors take this element into account when they evaluate the student, since they do not expect them to be so "bright" at the end of a two-hour exercise.
- the average workload since the beginning of the exercise has also an impact on the student's mental load.

To resume, the data recorded for each task (ex conflict resolution) is a set of performance measures (Pi) in various contexts (Ci), so a set of couples (Pi,Ci). Ci is already defined as (Si, Ri) and each situation Si is defined as ($\alpha$, $\varphi$, $\sigma$). Hence for a given task, like conflict resolution, we get for example the following data:

| Performance | Context (Ci) | | | | |
|---|---|---|---|---|---|
| | Situation (Si) | | | Resource load (Ri) | |
| | α | φ | σ | W | M |
| P1 | 8 | Climb/Cruise | Same perf | 5.64 | 4 |
| P2 | 12 | Cruise/Cruise | Quite different | 5.44 | 4 |
| P3 | 82 | Cruise/Cruise | Same perf | 6 | 4.6 |
| P4 | 98 | Descent/Cruise | Same perf | 6.90 | 5.2 |
| ..... | | | | | |

Fig 3. List of performances in context

The next section will describe the analysis of these performances in context.

## 5. Analysis of performance in context: towards a cognitive assessment of learning

By adding the contextual information to the performance measures and by reasoning on the situations and resource load associated to them, we are able to assess:

- the stability of performance across various situations.
- the level of automaticity level of his/her components

**Assessment of stability**

The level of stability can be measured by observing the changes in student's performance across a variety of situations at similar workload. It is particularly useful for instructional purposes. If the student performs consistently in any situation, then the workload can be increased. If not, then the analysis of the contextual information may help identifying the situation pattern(s) causing problems.

We use what we call a "context filter" to filter the contextual data with a condition on one of the components α, φ, σ or R.
For example if the condition is:

$$45^\circ < \alpha < 135^\circ$$

then it will filter all the situations involving "crossing track" type of conflicts.

We use this filter to group the initial set of data {(Pi,Ci)} under main categories of resource load. For each resource load category, we build a graph of performances against situations.

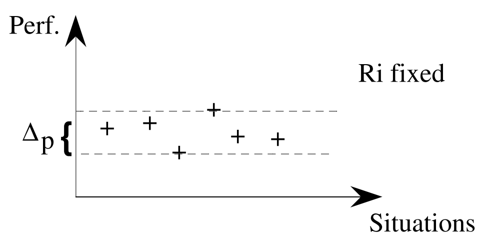


Fig 4- Performance chart across situations

Here we want to assess whether the performance is consistent, ie we need to measure the dispersion of the data. We will use the coefficient of variation CV and the size of the range of performance, $\Delta_p$.
If CV is important, then there is no stability. More practice is needed before increasing the resource load.

If CV is small and the $\Delta_p$ is large, then the data is concentrated around the average performance, but there are a few measures which are out of range. The situations corresponding to these measures need to be analysed further.
If CV and $\Delta_p$ are both small, then there is a good stability and the resource load may be increased.

**Assessment of automaticity**
The level of automaticity on a particular task can be measured by observing the impact of an increased resource load on the student's performances. When the student's performance on a task is consistent across all the situations (which is given by the previous assessment), the resource load will gradually be raised. If the student continues to perform well, even with a slight drop, then the level of automaticity is high.

The initial set of data {(Pi,Ci)} is sorted using (1) the resource load component Ri, (2) only the workload Wi or only the Mental load Mi. We then analyse the three graphs of performance as the load increases.

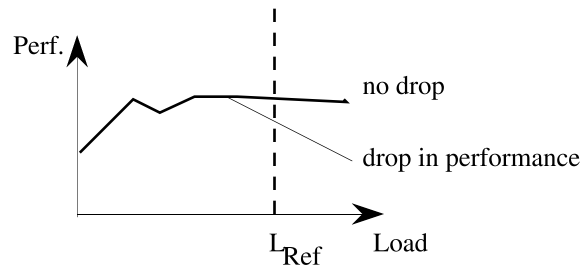


Fig 5 - Performance chart against load

The three graphs will provide information on how the student copes with a high resource load. The one against the workload says more about the complexity of the situations the student faces. A strong drop indicates that the student cannot perform all the tasks at the same time, especially when this drop is also observed in other tasks.
The chart against the mental load says more about the ability to concentrate for a long time at a high workload. A strong drop here indicates that the student is getting tired, has difficult concentrating, but does not mean that the knowledge is not acquired.

Since humans are not indefatigable, even experienced controllers have a limit above which their performance drops. This threshold ($L_{Ref}$) is pre-fixed by the instructor. The level of automaticity can considered to be acquired when there is no substantial drop in performance before this "human" limit.

## 6. Implications for the learner model in ATEEG

The learner model in the first version of our ATEEG system implements the following for the tasks of conflict detection and conflict resolution [Yacef & Alem 96b] :

- increments the number of "times seen" (S)

- uses a set of performance measures compared to expert measures to calculate a "performance level" (P): 100% being similar performance than the expert. The expert performance is generated by the Expert module. However, due to the nature of ATC activity, this module do not pretend to provide the best solution, only a good one according to general ATC standards.

For conflict detection these performance measures are about timeliness: time to detect a conflict, time of detection before estimated time of conflict.
For conflict resolution the performance measures are about (1) timeliness: time to resolve a conflict, time of resolution before estimated time of conflict and (2) quality: economy of the resolution strategy, additional workload generated (extra conflicts), etc..

We are currently implementing the next version of ATEEG system in which the learner model incorporates the assessment of the levels of stability and automaticity as presented in this paper.

## Conclusion

The work described in this paper is about the training of ATC operational skills. We have shown that for this type of non-algorithmic and dynamic activity, assessing the learning in terms of performance is not sufficient. Task performance data needs to be interpreted in context. The context in ATC skill training is composed of characteristics of the traffic situation and of resource load. Capturing such information allows a diagnostic of the learning at a cognitive level. Our current effort lies in improving our student model engine [Yacef & Alem 96b] in ATEEG in order to demonstrate the gain in terms of learning and instructional strategies.

## References

Alem, L. and Keeling R. (1996a) "Intelligent Simulation Environment an Application to Air Traffic Control Training". *In Proceeding of the first international conference on Advancing simulation technology and training*, SimTecT'96. March 96, Melbourne, pp 323-328.

Alem L., Woolf B.P., Munro A. & van Joolingen W. (1996b), Proceedings of Simulation-based learning technology workshop, held in conjunction with the Third International Conference in Intelligent Tutoring Systems (ITS'96), Montreal, June 96.

Anderson J.R. Boyle C.F & Yost G. (1985), The geometry tutor. In A.Joshi (Ed), *In Proceedings of the Ninth International Conference on Artificial Intelligence* (pp 1-7). Los Altos, CA: Morgan Kaufman.

Bisseret A. , Redon S. & Falzon P. (1981), Computer-aided training for risky decision making, INRIA Rocquencourt rapport de recherche #95, Octobre 1981.

Boy G.A., (1991), Intelligent Assistant Systems, Knowledge-Based Systems Vol 6. Academic Press, Harcourt Brace Jovanovich, Publishers. ISBN 0-12-121245-9.

Brown J.S, Burton R.R. & deKleer J. (1982), Pedagogical, natural language and knowledge engineering techniques in SOPHIE I, II and III. In D. sleeman & J.S. Brown (Eds.) *Intelligent Tutoring Systems*, New York: Academic Press, pp 227-282.

Cannat J.J., Ly S., Miailler B. (1987), Approche IA pour la modelisation du travail du controleur aerien, *In Proceedings of IA et circulation aerienne,* ENAC, 1987.

Chu, R. W., Mitchell, C. M. and Jones, P. M. (1995) Using the operator function model and OFMspert as the basis for an intelligent tutoring system, *In Proceedings of IEEE Transactions on Systems, Man and Cybernetics* Vol. 25(7): 054-1075.

Eliot, C. & Woolf, B (1994), Reasoning about the user within a simulation-based real-time training system, *Proceedings of 4th Int. conf. on User Modeling*, Hyannis, Massachusetts, pp121-126.

Gaillard I. (1992), Analyse de l'activite et des savoir-faire d'operateurs experts : le cas des controleurs du trafic aerien lors du changement de la position de controle", PhD thesis, University of Paris-Nord.

Hollan J.D., Hutchins E.L. & Weitzman L. (1984), STEAMER: an interactive inspectable simulation-based training system, *The AI Magazine*, Summer 1984, pp 15-27.

Rasmussen J. (1983). Skills, rules and knowledge: signals, signs, and symbols, and other distinctions in human performacne models. *In IEEE transactions on Systems, Man and Cybernetics*, 3, pp257-266.

Regian, W. & Schneider, W. (1990), Assessment procedures for predicting and optimizing skill acquisition in extended training. In Fredericksen (Ed.), *Diagnostic monitoring of skill and knowledge acquisition*. Hillsdale, NJ : Erlbaum.

Reiser B.J., Anderson J.R. & Farrel R.B. (1985), Dynamic student modelling in an intelligent tutorfor LISP programming, *Proceedings of IJCAI'85*, vol 1, pp 8-14, Los Altos, CA: Morgan Kaufman.

Schneider W. (1988), Developing automatic component skills for high workload performance in ATC tasks, *Symposium on ATC Training for tomorrow's technology*, Dec 88, Oklahoma, US.

Suchman L.A. (1987) Plans and situated actions: th eproblem of human-machine communication. Cambridge: Cambridge University Press.

White B.Y. & Frederiksen J.R. (1987), Qualitative models and intelligent learning environments, *Artificial intelligence and Education*, Vol.1, (pp. 281-305). Norwood, NJ: Ablex, W. Lawler & M. Yazdani (Eds).

Woods, D.D. (1988), Coping with complexity: the psychlogy of human behavior in complex systems, *Tasks, errors and mental models*, Goodstein, Andersen and Olsen eds. London: Taylor & Francis, pp 128-148.

Yacef, K & Alem, L. (1996a) Student and expert modelling for simulation-based training : a cost effective framework. *In Proceedings of Third International Conference in Intelligent Tutoring Systems (ITS'96)*, Montreal, June 96.

Yacef, K & Alem, L. (1996b) Evaluation of learner's skills in the context of dynamic and complex systems. *In Proceedings of IEEE Systems, Man and Cybernetics (SMC'96)* , Beijing, China, October 96.