2005

Competence-Based Assessment Design for Air Traffic Control Training

Esther Oprins

Ernst Burggraaff

Hans van Weerdenburg

COMPETENCE-BASED ASSESSMENT DESIGN FOR AIR TRAFFIC CONTROL TRAINING

Esther Oprins, **Ernst Burggraaff** & **Hans van Weerdenburg**
Air Traffic Control the Netherlands (LVNL)
Amsterdam, The Netherlands

Air Traffic Control (ATC) simulator and on-the-job training (OJT) requires a valid and reliable assessment system. Competence-based assessment results in more effective learning processes, better pass/fail decisions and improved selection criteria which may contribute to an increased output of competent controllers from training. This paper describes the design of the assessment system in use by Air Traffic Control the Netherlands (LVNL).

Introduction

Among other process control tasks in transportation and process industries, the ATC task is called a complex cognitive skill (Van Merrienboer, 1997). The processing of large amounts of dynamically changing information calls for complex cognitive processes. In combination with the strict safety requirements that do not allow any (human) error, this puts high demands on controller's competences. Selection requirements are generally very high, but often the outcome of training remains still too low, especially in ATC organizations serving busy and complex airports such as Schiphol Airport (LVNL). This may result in a shortage of controllers. Besides, high failure rates are undesirable because of the time-consuming and expensive simulator training and OJT. Solutions can be sought in improved selection, training, or both. An important part of training design is assessment. A valid and reliable assessment system may contribute to an increased output of competent controllers in several ways. First, assessment is a base for adequate feedback which supports trainee's learning processes. Second, training can be adapted to the trainee's needs which may increase learnability for more candidates. Third, more accurate judgments lead to better founded pass/fail decisions and help to reduce false positives and false negatives during later training phases. Fourth, higher reliability of training criteria makes it possible to obtain higher predictive validities for selection instruments.

Unfortunately, despite of its importance, scientific research on assessment in ATC training seems to have been very limited. The scarce literature is restricted to descriptions of ATC assessment from a more practical perspective (e.g. Hopkins, 1995). Some studies on ATC selection are relevant, because they involve assessment in work samples or criterion development for validation (e.g. Ramos, Heil & Manning, 2001). Within the field of aviation a substantial amount of research has been done on assessment of aircrew (e.g. O'Connor et al, 2002). The ATC task, however, is different. Its complex and time-critical character makes assessment extremely difficult to design. The invisible, cognitive processes result in subjective judgements of assessors, who mostly depend on over-the-shoulder observation.

The aim of this paper is to describe the design of the assessment system in ATC training at LVNL. We discuss how assessment based on competences may lead to more effective learning processes, better pass/fail decisions and improved selection criteria.

Competence-based Assessment: Background

The assessment system is based on principles of competence-based assessment found in the literature.

Competences

We define competence in relation to training as follows: 'the ability to apply acquired knowledge, skills and attitudes while performing tasks in realistic settings'. Competences are the result of a learning process for which a person needs specific abilities, personality and other features included in selection requirements (Roe, in press). Competences become implicit by learning since information processing has been automated. Although competences are not innate, they differ in trainability. Schneider (1990) separates 'consistent' components that improve by practicing from 'non-consistent' components that do not necessarily improve. This relates to 'recurrent' and 'non-recurrent' skills (Van Merriënboer, 1997).

Assessment of Competences

Assessment in ATC simulator training and OJT is usually defined as 'performance assessment' (Wickens, Mavor & McGee, 1997). An assessor judges performance criteria on a rating scale on the base of over-the-shoulder observation. These criteria are generally formulated in observable behavior, also called 'behavioral markers' (O'Connor et al, 2002). We consider competence-based assessment to be a specific type of performance assessment due to its focus on competences. In accordance with modern learning theories (Pellegrino, Chudowski & Glaser, 2001), competences are not analytically split up in detailed skills and knowledge, but assessment takes place at a higher level. This allows for different learning curves: (sub)skills and pieces of knowledge may be learned in a different order or tempo, as long as the competences required are obtained after a certain (flexible) learning period. Further, due to their more generic character, the same competences can be assessed during training. For instance, planning is relevant in each ATC training phase as well as in each ATC task execution. This makes it easier to identify trainee's strengths and weaknesses and to define appropriate training interventions at an early stage. In addition, progression on each competence can be measured, providing an important indicator of whether a trainee is still learning. This is essential in pass/fail decisions: when a trainee has not reached a learning plateau yet, it may be useful to continue training. In this respect, trainability of ATC competences must be taken into account (Schneider, 1990): consistent components (e.g. radiotelephony) that are not mastered yet could still be improved in contrast with non-consistent components (e.g. planning) that are more often reasons for failing.

In order to get a complete picture, covering the cognitive, emotional and social aspects, assessment should involve all these aspects that belong to a competence. In assessment of aircrew these are referred to as 'non-technical skills' (O' Connor et al, 2002). Many of them, such as situational awareness and decision making, are also essential for ATC.

A crucial step in the design process is a competence analysis. Involvement of air traffic controllers as subject matter experts (SME's) is extremely important because their implicit knowledge has to be explicated as the reverse process of learning.

Cognitive Processes

The assessment of cognitive processes is extremely important in ATC. This calls for an inference from observable behaviour and interaction with the trainee (e.g. asking questions). Feedback is more effective when coaches have insight into trainee's thinking patterns and strategies, which needs more emphasis in ATC training (Schneider, 1990). Besides, assessment of cognitive processes is required to obtain diagnostic information on performance shortcomings and to predict future performance in ATC training (Regian & Schneider, 1990).

The importance of assessment of cognitive processes is one reason why 'automated measurement' in ATC simulator training has hardly been applied, although safety and efficiency aspects such as separation, conflicts, delay and communication can be logged by the computer (Wickens, Mavor & McGee, 1997).

Selection

Competence-based assessment may indirectly contribute to better training results by using the competences as criteria in selection, because they optimally reflect the personal basis of job performance (Roe, in press). Performance measures obtained in training must be reliable and valid to make validation research valuable. Further, in work samples, as part of selection systems, competences can be rated using similar performance criteria as in training, serving as predictors for future performance. Thus, many similarities can be found between performance criteria applied in ATC training and in work samples (e.g. Ramos, Heil & Manning, 2001).

Psychometric Requirements

A precondition for any assessment system is its psychometric quality. Reliability and validity can be obtained by judgments of multiple assessors, assessor training, sophisticated measurement techniques (e.g. 'behavioral anchored scales'), representative tasks and performance criteria and so on (Berk, 1986).

Design Method

A competence analysis and literature research resulted in the ATC Performance Model which has served as a framework for the assessment design.

Competence Analysis

We organized two competence workshops in which twelve air traffic controllers formulated a set of thirteen competences. The set consists of: situational awareness, decisiveness, dealing with unexpected situations, workload management, conflict solving, multitasking, prioritizing, co-ordination and communication, flexible planning, leadership, teamwork ability, perseverance, and critical attitude. Each competence is supported by a set of eight to twelve behavioral markers. The collaboration tool Meetingworks was used, which makes it possible to brainstorm, discuss and structure electronically. This method enabled controllers to come to agreement about the completeness and the interpretation of each competence with aid of the behavioral markers, formulating them in their own jargon. This makes the competences recognizable and practically usable in training. Controllers were forced to think about their own work performance at a more abstract level.

Besides, we did literature research looking for additional aspects of ATC performance that might have been forgotten in the workshops. We were also interested in the relations between these aspects in order to categorize the thirteen competences. Thus, we compared our set with existing ATC (cognitive) task analyses (e.g. EATMP, 1999), performance models (e.g. Hadley, Guttman & Stringer, 1999), and performance measurement systems (e.g. Ramos, Heil & Manning, 2001). On the basis of this we developed the ATC Performance Model (Oprins & Schuver, 2003), presented in figure 1.

The ATC Performance Model

The model shows the dominant role of information processing in ATC work. Information processing provides the basis for actions, which result in the outcome: handling of air traffic. The way in which this happens depends on a number of influencing factors. All components of the model are specified in terms of competences. We recognize the majority of the competences defined in the workshops in the dark gray parts, some in the white parts. Some were revised and others added as a result of the additional literature research. We see that information processing comprises *situation assessment, planning* and *decision making.* This is mainly derived from the ATC model of Hadley, Guttman and Stringer (1999), but these cognitive processes are not necessarily ATC-specific in contrast with the actions and outcome. Situation assessment (e.g. Endsley, 1995) is further divided into *perception*, *attention management* and *interpretation* (mental picture). The actions consist of *communication, co-ordination, strip/label management,* and *equipment operation.* The outcome distinguishes *safety* and *efficiency.* The influencing factors are mainly represented by *workload management* and *teamwork ability*.

Properties of the Assessment System

Performance Criteria

Setting new performance criteria was the most fundamental change in the previous assessment system. They are directly derived from the ATC Performance Model. Each criterion is rated on a 6-points rating scale that strictly separates sufficient from insufficient behavior. A set of related criteria, formulated in terms of observable behavior, form a category representing a specific competence. Each category is visible in the model as a dark gray part and is marked in italics. The typical Dutch jargon proposed in the workshops has been maintained in order to maximize recognition and comprehension by the users. Most criteria are literally identical to the behavioral markers formulated in the workshops. The same fourteen categories are used for all ATC functions (e.g. area, aerodrome control), from the start of initial training till final job performance. They are even applied in two work samples that are part of the LVNL selection system. The criteria are also identical for each ATC function when possible, for instance, criteria of the category *communication*:

Communication ***(all ATC functions)***
- Applies (non) standard phraseology correctly
- Express himself clearly, unambiguously and shortly
- Provides correct and sufficient traffic information

Only the criteria that belong to *safety* and *efficiency* are different because they have another meaning in each ATC function, illustrated by next example:

Safety ***(ground control)***
- Prevents runway intrusions
- Arranges conflicts and right-of-way situations
- Checks correctness of clearances on strips and EDD
Safety ***(area control)***
- Maintains separation minima correctly
- Builds in sufficient safety buffers
- Switches from monitoring to vectoring in time

The use of the same categories and performance criteria makes it possible to follow trainee's progression on each competence in order to define appropriate training interventions, based on trainee's weaknesses. The criteria can be applied in different task situations which provides a complete picture about trainee's performance. They are independent of variables such as traffic complexity or specific events which are relevant for the design of assessment tasks.

The ATC Performance Model provides indications on how the performance criteria can be assessed. First, the model separates objectively measurable criteria (outcome, actions) from criteria that can only be assessed subjectively (information processing). We have argued that safety and efficiency could even be 'automatically' measured in the simulator. This distinction is useful for assessors who have to be aware of their own restrictions when they depend on subjective measures. Second, the model gives information about trainability. Actions are trainable because they improve by practicing in contrast with the majority of the cognitive processes. The latter express the 'gut feeling' of assessors. They help them to argue why trainees perform (in)sufficiently as causes for (in)sufficient actions or outcome, relevant for adequate feedback and pass/fail decisions.

Phasing

We divided the training period into phases and determined performance standards to be achieved at the end of each phase. Trainee's competence is assessed against lower standards in earlier phases. Before the introduction of phasing assessors did only rely on their experience, which increased disagreement between them. Trainees did not have insight into the standards required in final or in intermediate phases. This vagueness did certainly not contribute to learning and to succeed in training.

In simulator training, phases are mainly defined by the sequence of simulator scenarios (Farmer et al, 1999). OJT normally occurs 'unstructured', not only in ATC (Jacobs & Jones, 1995). Structuring OJT is difficult because assessment tasks cannot be arranged due to the ongoing live traffic. We divided OJT in phases based on three basic principles: degree of *safety/efficiency*, *complexity* of traffic situations, and *independence* of the coach. Each OJT consists of four phases with flexible lengths, dependent on trainee's assessment results (progression), to take into account individual differences in learning. An example of OJT phases in area control is the following:

Phase 2 (8-14 weeks): to be able to handle *standard* traffic both *safely* and *efficiently*, and *complex* traffic *safely*, independently of the coach.
Phase 3 (8-14 weeks): to be able to handle both *standard* and *complex* traffic both *safely* and *efficiently* independently of the coach

Standard and complex traffic are further detailed in terms of traffic intensity, diversity in aircraft, flight destinations, weather circumstances, runways in use, specific events and so on. These variables are predetermined in simulator scenarios (assessment tasks), but in OJT only a description for trainees and assessors is available serving as a guideline. Safety and efficiency, as well as the other competences, are specified for each separate ATC function and for each phase in both simulator training and OJT. Therefore the performance criteria are accompanied by behavioral examples, illustrated in figure 2, which can be considered as a variant on 'behaviorally anchored scales' (Berk, 1986). The examples do not specify the scale positions, but represent the performance standards to be achieved at the end of each phase. Differences between ATC functions, also for the criteria that stay the same (e.g. planning, communication), become directly visible in these examples, which are necessarily function-specific to be as clear as possible. The examples contribute to consistent judgments between assessors, not only for assessing against the same performance standards in a phase but also for assigning specific behavior to the same criterion. For trainees it is clearer what is expected from them in a specific training phase.

Continuous Assessment

Continuous assessment is applied as in the majority of ATC organizations (Hopkins, 1995). Coaches, who are also assessor, measure trainee's achievement during training. They are continuously in interaction with the trainee and can force trainees to verbalize their thoughts. This enables them to assess cognitive processes (e.g. strategies). Representativeness of task situations is guaranteed, because assessment is not restricted to a particular moment. Multiple assessors are involved for maximizing reliability, who are trained beforehand in the use of the system, interpretation of performance criteria, and avoidance of rating errors. However, objective measurement is impossible since coaches are constantly influencing trainee's task performance by their guidance. Therefore, we combine continuous assessment with performance tests (Berk, 1986).

Performance Tests

Performance tests measure trainee's performance during a test in the simulator or in OJT without coaching interventions. We emphasize the objective character of the test as a counterpart of continuous assessment in several ways. In the simulator checklists are used for the observation of events occurring in scenarios at a specific time (e.g. conflicts, runway changes), added by possible solutions for each event. The solution chosen by the trainee is marked. Afterwards the final test score is calculated. This final score is the sum of weighted scores that are assigned to each criterion. The weighting relates to the ATC Performance Model: the sum of the scores belonging to information processing has the same weight as the actions and outcome together, because information processing refers to the causes for (in)sufficient actions and outcome. Safety is measured objectively by counting the number of safety violations (e.g. unsolved conflicts), based on the annotations on the checklist.

Conclusions and Future Directions

The assessment system has been used in practice for two years now. The evaluation of the system comprises several parts. First, we investigated the practical use of the system and the improvement of learning processes for coaches and trainees qualitatively (interviews, questionnaire, report analysis). This has led to positive results. The involvement of controllers in the design process has contributed to a better recognition of behavior. Assessors more easily express and validate their ‘gut feeling’ with aid of the competences, which results in more appropriate training interventions and better founded pass/fail decisions. The ATC Performance Model helps them to get insight into the different components of performance, such as the distinction between objective and subjective measures and the extent of trainability of competences. The use of the same performance criteria makes assessors more familiarized with their meaning. It also helps them to follow trainee’s progression on each competence and to provide adequate feedback. Agreement about performance standards in different phases is higher. For trainees it is clearer which competences they must develop further in a specific phase. Thus, the assessment system is definitely practically usable and contributes to more effective learning processes.

Second, we are investigating the psychometric quality of the system, especially the interrater reliability, internal consistency, and predictive validity. However, more long-term evaluation is needed for quantitative conclusions about a possible increased output from training, although the findings about improved learning processes are encouraging. This evaluation research has to be regarded in relation to the selection system, which we have redesigned simultaneously using the competences as criteria, and other possible influences (e.g. changes in training design). This makes it all rather complex. For facilitation of this further research we make use of a database that stores all selection and training results. Therefore, assessors fill in assessments digitally by means of the web-based assessment tool Questionmark Perception. This tool has several advantages, not only for research purposes. First, trainee’s progress can be better followed by interested persons (e.g. training managers) who have access from several places so that interventions can be undertaken as soon as possible. For instance, from the office there will be direct access to assessments that takes place at the tower. Second, different overviews and graphs (e.g. individual learning curves) can be distillated from the system, which provides more insight in learning processes. Third, reliability is increased because the system forces assessors to fill in assessment reports accurately and univocally. Finally, training results can more easily be used for validation studies for both selection and training purposes, needed for long-term evaluation. This research will be an on-going process which makes it possible to adapt performance standards in training and selection requirements continuously in order to maximize output from training ultimately.

References

Berk, R.A. (1986). *Performance assessment: methods and applications*. The Johns Hopkins University: Baltimore.

EATMP (1999). *Integrated task and job analysis of air traffic controllers phase 2: task analysis of en-route controllers* (HUM.ET1.ST01.1000-REP-04). Luxemburg: Eurocontrol.

Endsley, M. (1995). Towards a theory of situational awareness in dynamic systems. *Human factors, 37*, 32-64.

Farmer, E., Rooij, J. van, Riemersma, J., Jorna, P., & Moraal, J. (1999). *Handbook of simulator-based training*. Aldershot: Ashgate.

Hadley, G., Guttman, J., & Stringer, P. (1999). *Air traffic control specialist performance measurement database* (DOT/FAA/CT-TN99/17). Washington DC: FAA.

Hopkins, D. (1995). *Human factors in air traffic control*. London: Taylor & Francis.

Jacobs, R.L., & Jones, M.J. (1995). *Structured on-the-job training. Unleashing employee expertise in the workplace*. San Francisco: Berrett-Koehler Publishers.

Merrienboer, J.J.G. van (1997). *Training complex skills: a four component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.

O’Connor, P., Hormann, H., Flin, R., Lodge, M., & Goeters, K. (2002). Developing a method for evaluating crew resource management skills: a European perspective. *The international journal of aviation psychology, 12,* 263-285.

Oprins, E., & Schuver, M. (2003). Competentiegericht opleiden en beoordelen bij LVNL. *HUFAG Nieuwsbrief, 6,* 2-4.

Pellegrino, J.M., Chudowski, N., & Glaser, R. (2001). *Knowing what students know. The science and design of educational assessment*. Washington: National Academic Press.

Ramos, R.A., Heil, M.C., & Manning, C.A. (2001). *Documentation of validity for the AT-SAT Computerized Test Battery: volume II* (DOT/FAA/ AM-01/6). Washington, DC: FAA.

Regian, J.W., & Schneider, W. (1990). Assessment procedures for predicting and optimizing skill acquisition after extensive practice. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.). *Diagnostic monitoring of skill and knowledge acquisition* (pp. 297-323). Hillsdale, NJ: Lawrence Erlbaum Associates.

Roe, R.A. (in press). The design of selection systems: contexts, principles, issues. In A. Evers, O. Smit, & N. Anderson (Eds.), *Handbook of personnel selection*. Oxford: Blackwell.

Schneider, W. (1990). Training high-performance skills: fallacies and guidelines. In M. Venturino (Ed.), *Selected readings in human factors* (pp. 297-311). Santa Monica, CA: The human factors society.

Wickens, C.D., Mavor, A.S., & McGee, J.P. (1997). *Flight to the future: human factors in air traffic control*. Washington: National Academy Press.

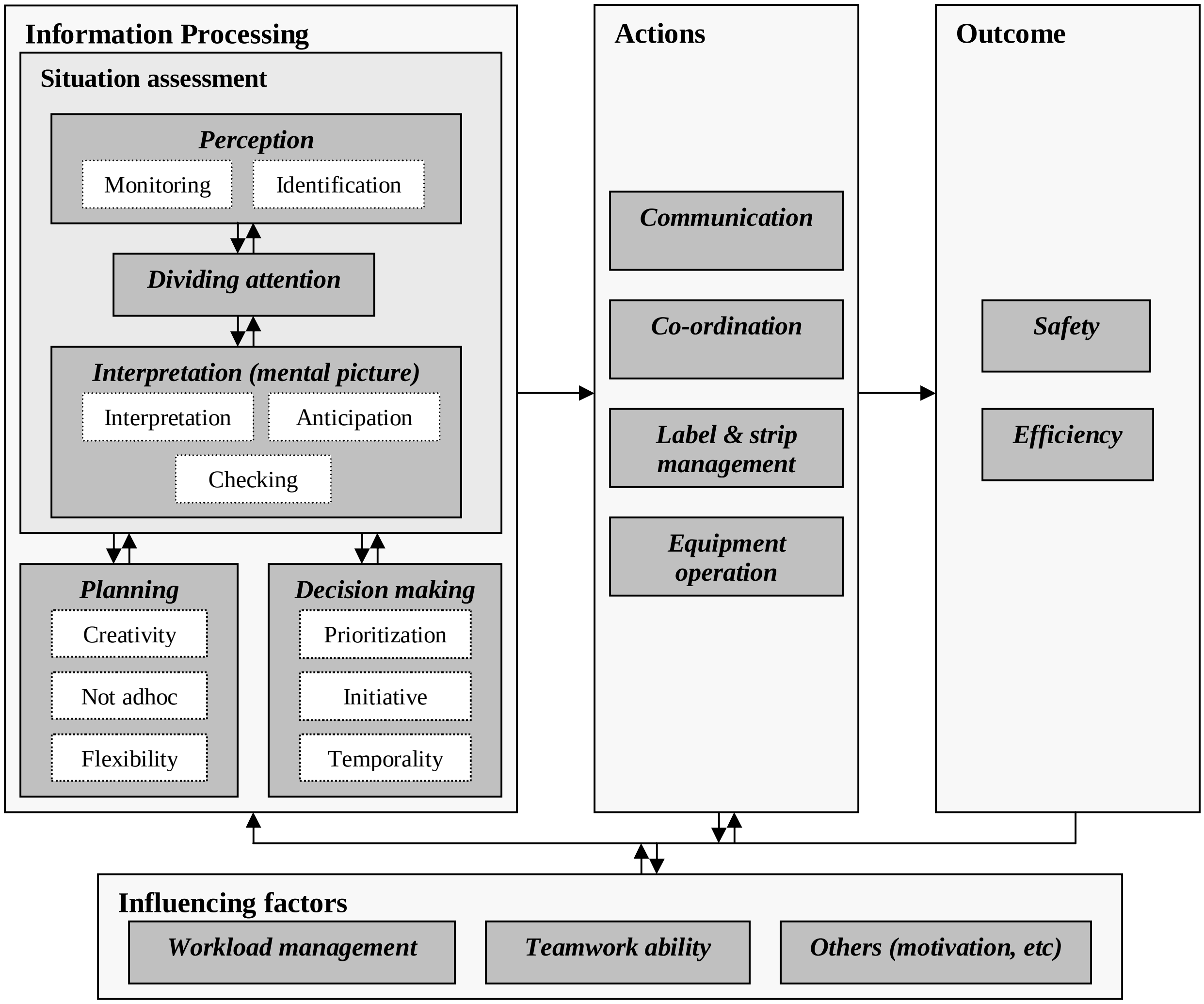

Figure 1. *The ATC Performance Model (Oprins & Schuver, 2003)*

Efficiency	**Phase 1**	**Phase 2**
• Applies speed control correctly • Applies vector technique correctly • Takes into account aircraft performances • Takes into account different flight levels • Builds a sequence of climbing and descending traffic	Speed control and vector technique does not need to be optimal, but must be conflict free and conform standard routes and transfer, taking into account differences in aircraft performances and in time turning to own or published navigation or speed. Some delay may still occur; sequences are not always efficiently enough yet. Application of level separation and assessment of intermediate levels during sequencing is not always optimal, for instance, 2 flights in different STARS are cleared to the same FL.	Sequences consist of 5 to 7 NM interval, with minimal speed differences by optimal speed control of inbounds, and with parallel handling over of outbounds by optimal vectoring. There is a striving for continuous climb/descend, taking into account differences in aircraft performances and in time turning to own or published navigation or speed. Delay has been avoided whenever is possible. Level separation and assessment of intermediate flight levels during sequencing is applied optimally.

Figure 2. *Performance standards in two phases of area control training for Efficiency.*