EUROPEAN ORGANISATION
FOR THE SAFETY OF AIR NAVIGATION

EUROCONTROL

This document is issued as an EATMP Guideline. The contents are not mandatory. They provide information and explanation or may indicate best practice.

Selection Tests, Interviews and Assessment Centres for *Ab Initio* Trainee Controllers: Guidelines for Implementation (Revised)

Edition Number	**:**	**2.0**
Edition Date	**:**	**17.09.2002**
Status	**:**	**Released Issue**
Intended for	**:**	**EATMP Stakeholders**

DOCUMENT CHARACTERISTICS

TITLE

Selection Tests, Interviews and Assessment Centres for *Ab Initio* Trainee Controllers: Guidelines for Implementation (Revised)

EATMP Infocentre Reference:	020820-01
Document Identifier	**Edition Number:** 2.0
HRS/MSP-002-GUI-03-01	**Edition Date:** 17.09.2002

Abstract

This document addresses guiding principles and gives policy recommendations in relevant areas in recruitment and selection of *ab initio* trainee controllers and in related areas. The main focus is on implementation, quality assurance and maintenance of appropriate selection tools and methods. Administration and professional practices in recruiting, selection and training are addressed with a view to help to introduce new or to adjust existing policies. The aim is higher effectiveness (validity), efficiency in the use of methods and tools throughout the whole recruitment, selection and training process. In addition the guidelines describe professional practices in regard to quality assurance of the recruiting, selection and training system that will help to target potential problems or shortcomings in an existing system or to implement new methods. This is done with the aim to maintain and improve the quality and validity of used methods and tools.

Keywords

Ability test	*Ab Initio* Trainee	Achievement test	Assessment centre
Cut-off score	Interview	Recruitment	Reliability
Selection	Task analysis	Validation	Validity

Contact Person	Tel	Unit
H. Rathje	+32 2 729 3955	Human Factors and Manpower Unit (DIS/HUM)

STATUS, AUDIENCE AND ACCESSIBILITY

Status		Intended for		Accessible via	
Working Draft	☐	General Public	☐	Intranet	☐
Draft	☐	EATMP Stakeholders	☑	Extranet	☐
Proposed Issue	☐	Restricted Audience	☐	Internet (www.eurocontrol.int)	☑
Released Issue	☑	*Printed & electronic copies of the document can be obtained from the EATMP Infocentre (see page iii)*			

ELECTRONIC SOURCE

Path:	G:\Manpower Sub Programme\Deliverables	
Host System	**Software**	**Size**
Windows_NT	Microsoft Word 8.0b	663 Kb

EATMP Infocentre
EUROCONTROL Headquarters
Rue de la Fusée, 96
B-1130 BRUSSELS

Tel: +32 2 729 5151
Fax: +32 2 729 9984
E-mail: eatmp.infocentre@eurocontrol.int

Open on 08:00 -15:00 UTC from Monday to Thursday, incl.

DOCUMENT APPROVAL

The following table identifies all management authorities who have successively approved the present issue of this document.

AUTHORITY	NAME AND SIGNATURE	DATE
Please make sure that the EATMP Infocentre Reference is present on page ii.		
Manager HRS Manpower Sub-Programme (MSP) Chairman Review Group	H. RATHJE	11.09.2002
Manager EATMP Human Resources Programme (HRS-PM)	M. BARBARINO	11.09.2002
Chairman EATMP Human Resources Team (HRT)	A. SKONIEZKI	12.09.2002
Senior Director Principal EATMP Directorate (SDE)	W. PHILIPP	16.09.2002

DOCUMENT CHANGE RECORD

The following table records the complete history of the successive editions of the present document.

EDITION NUMBER	EDITION DATE	INFOCENTRE REFERENCE	REASON FOR CHANGE	PAGES AFFECTED
0.1	22.07.1997		Working Draft	All
0.2	01.09.1997		Draft	All
0.3	18.09.1997		Proposed Issue	All
1.0	27.02.1998		Released Issue	All
1.1	18.06.2001		Draft reviewed by Selection Experts User Group'	All
1.2	19.11.2001		Draft approved by HRT Manpower Sub-Group (MSG) as Proposed Issue for submission to HRT17	All
1.3	08.02.2002		Document configuration adjustments (structure, layout, styles) to Proposed Issue	All
2.0	17.09.2002	020820-01	Released Issue (inclusion of Chapter 15 on 19.08.2002 + last document configuration adjustments)	All

CONTENTS

EXECUTIVE SUMMARY

This document was part of the work done in 1996-97 by the Selection Task Force (STF) II of the Human Resources Team (HRT) of the European Air Traffic Harmonisation and Integration Programme (EATCHIP), now known as the European Air Traffic Management Programme (EATMP).

This document addresses guiding principles and gives policy recommendations in relevant areas in recruitment and selection of *ab initio* trainee controllers and in related areas. The main focus is on implementation, quality assurance and maintenance of appropriate selection tools and methods. Administration and professional practices in recruiting, selection and training are addressed with a view to facilitate the introduction of new or the adjustment of existing policies.

Such guidelines can only do so much. If, in their final version, the guidelines are too vague, they will make little difference. If, on the other hand, they are too rigid, they will be impossible to apply to the varied national administrations or Air Traffic Services Provider (ATSP) organisations. The burden of balancing these was of major concern to the STF. Necessary cautions were given in this document whenever desirable goals were likely to go beyond reasonable expectations as judged by many ATSP organisations.

This document intends to give decision makers in selection, managers responsible for implementing, administering or providing the selection service, selection experts, specialists and professional psychological audiences working in this field detailed information on implementation of selection tools and methods for *ab initio* trainee controllers. It will be presented at a level which would enable a wide range of persons to use it.

The aim of this document is higher efficiency in the use of methods and tools throughout the whole recruitment, selection and training process. In addition, the guidelines describe professional practices in regard to quality assurance of the recruitment, selection and training system, which will help to target potential problems or shortcomings in an existing system or to implement new methods. This is done with the aim to maintain and improve the quality and validity of methods and tools in use.

Several examples are presented using Air Traffic Control (ATC) procedures along with other kinds of data to illustrate how professional practices in test construction, validation, interviewing or assessment might be developed. Several cautions are extended regarding the use of some statistical procedures.

Due regard should be given to the Technical Supplement (TS) of this guidelines document (EATCHIP, 1998a), hereafter referred to as the TS. Cross-references between the two documents are provided whenever relevant. It is recommended that this guidelines document will be read simultaneously with the TS.

A group of selection experts (Selection User Group) has reviewed the original document in 2001 to identify areas in need of update / revision and amendments. The Group based its considerations on findings in a guidelines implementation survey and with a view towards new developments of the Selection Core Drafting Group (CDG). The CDG was established

by the Manpower Sub-Group (MSG) as the Expert Working Group in the Human Resources Programme (HRS), Manpower Sub-Programme (MSP), for the Work Package "First European ATCO Selection Test (FEAST)" (HRS/MSP-002).

This document reflects the changes adopted by the Group.

Chapter 1, "Introduction", describes the background, purpose and scope of the document.

Chapter 2, "Task and Ability Analysis of Controllers Job", provides guiding principles on Task Analysis (TA) and Integrated Task and Job Analysis.

Chapter 3, "Training and Assessment Issues", addresses training issues linked to selection and to validation of selection tools.

Chapter 4, "Validation Issues", gives a general introduction to methodological issues which are detailed in the TS. This Chapter also describes the utility of validation and outlines some validation strategies.

Chapter 5, "Framework for Implementing a Recruitment and Selection Process", tackles in detail the mission, objectives, goals, policies and practical implementation strategies of a recruitment and selection process.

Chapter 6, "Recruitment Issues", highlights briefly some of the influencing factors and some of the ways in which external recruitment of *ab initio* trainee controllers can be carried out.

Chapter 7, "Selection Tests", addresses the underlying notions of selection tests and the considerations that should guide the implementation and practical application of these tools.

Chapter 8, "Selection Interview", outlines the assumptions and principles in the development of best practice in conducting selection interviews with recommendations for practical application.

Chapter 9, "Assessment Centre in Selection", outlines the assumptions and principles in the development of Assessment Centre (AC) and gives practical recommendations for its application.

Chapter 10, "Professional Practices in Test Application and Evaluation", gives consideration to several critical issues in the application and evaluation of selection tests.

Chapter 11, "Decision-making Procedures", outlines the underlying notions and gives recommendations with regard to decision-making in selection.

Chapter 12, "Recruitment, Selection and Training Database", outlines the principles and requirements of a standard database for *ab initio* trainee controller applicants, trainees and licensed air traffic controllers.

Chapter 13, "Future Trends and Steps", provides some realistic observations on events that are likely to happen in the future and which may affect future test developments.

Chapter 14, “Summary of Guidelines and Recommendations”, summarises the guidelines.

Chapter 15, “Another Approach in Selection, Recruitment and Training”, describes the approach taken in France to recruit Air Traffic Controllers.

The Annexes consist of References, a Glossary of Terms used in the document, a list of Abbreviations and Acronyms, and lists of the Contributors to this document and of the Review Group Members.

Page intentionally left blank

1. INTRODUCTION

At its fifth meeting the Human Resources Team (HRT), within the European Air Traffic Harmonisation and Integration Programme (EATCHIP), established the second Selection Task Force (STFII), which, under the auspices of the HRT, should provide further material for the introduction and application of specific selection methods in European Civil Aviation Conference (ECAC) States.

The Terms of Reference (TORs) of the Task Force were as follows:

- to collect information on available and emerging selection tools and methodologies for the selection of *ab initio* trainee controllers;
- to consider criteria for the evaluation of selection tools and methodologies;
- to consider the scope and content of a basic selection test battery for use in ECAC States;
- to consider the scope and content of guidelines on selection tests, interview and assessment centre.

The main task of the STF was to define evaluation criteria for a basic selection test battery for the selection of Air Traffic Controllers (ATCOs). Based on this an information survey was done worldwide to gather information on available and emerging selection tools that are currently in use and to report on the status. EATCHIP (1997) gives the information on tests and methods surveyed in 1996-97. A complete update of this information, done under the auspices of the Selection Core Drafting Group (CDG) and based on a survey in 2000 is available (see EATMP, 2000a).

The work envisaged in the third bullet above has now been taken up by the CDG in the frame of the Human Resources Programme, Manpower Sub-Programme (HRS/MSP) in Work Package "First European ATCO Selection Test package (FEAST)" (see EATMP, 2000b).

1.1 Purpose

The primary purpose of this document is to provide guidance material to national administrations for the implementation and application of proven state-of-the-art selection methods that should be used in the selection of *ab initio* trainee controllers.

Administration and professional practices in recruitment, selection and training are addressed with a view to help to introduce new or to adjust existing policies.

The aims are

i. Higher effectiveness and efficiency in the use of practices, methods and tools throughout the whole recruitment, selection and training system.

ii. To describe professional practices that help to identify problems or shortcomings in existing systems or to implement new methods.

iii. To continuously improve the quality and validity of the methods and tools used.

The main users of this document will be the persons responsible for the implementation and maintenance of selection tools for *ab initio* trainee controllers. Other users will include different experts and managers dealing with selection issues.

The document should be used as a source of information by persons involved in recruitment, selection and training, as well as Manpower Planning (MP).

1.2 Scope

The STF recognised the importance of good selection tests for many nations in the ECAC area. The document is based on the practical experience and professional practice available at present.

Further to this the STF recognised that the definition of a selection test battery is a complex, multi-step and iterative process.

The HRT Manpower Sub-Group (MSG) has tasked the CDG with this work package, to 'define, identify, create and provide a European controller test package which reflects future impacts and enables customised implementation and application' (see EATMP, 2000b).

For various reasons the STF has not tried to propose any one ‘definite solution’ for selecting ATCOs.

It was felt necessary by the STF that attention should be drawn into areas not directly within the TORs but which were nevertheless seen as directly affecting selection issues.

The following areas came under this category and were taken into consideration:

- task- and ability-oriented job analysis,
- recruitment policies and strategies,
- training and instruction issues,

- assessment of trainee performance,
- other criterion-related issues.

In addition to the guidance given in regard to the implementation and application of proven selection methods, serious consideration was given to:

- selection and training of selection personnel,
- methods for the pre-selection of candidates on the basis of application forms,
- validation and maintenance of selection test methods,
- recruiting, selection and training database requirements,
- methods for composite score derivation in selection,
- methods for deriving cut-off scores for decision-making.

The group members considered it necessary to recommend those practices in recruitment, selection and training which, if adopted, would not only assist in the implementation and application of enhanced selection methods but would also lead to:

- higher and comparable levels of abilities, knowledge and skills of trainees,
- higher success rates during training and higher operational proficiency.

To achieve this, the STF focused foremost on quality aspects in selection methods. However, practical aspects in implementation and application as well as costs were also taken into consideration.

Convergence and Implementation Programme (CIP) Objective 6.3.1 on Selection, agreed at the fifth meeting of the HRT, emphasises this opinion. It states that the basic requirements are:

- "Tests of proven objectivity, reliability and validity should be available to assess the relevant abilities, knowledge and the personality of the applicant."
- "Methods and procedures of test validation and control of the selection process should be in place."

EATMP conducted in 2000 a survey to identify areas where Air Traffic Service Provider (ATSP) organisations face difficulties in implementing EATCHIP selection guidelines. The survey found some challenges in defining validation performance criteria and in interviews' validation. The survey also showed the need to provide a proven, valid and reliable test package (FEAST) for (some) States.

The new European Convergence and Implementation Plan (ECIP) (EATMP, 2001a) defines under Objective HUM02: “Improve the recruitment, training and development of staff in order to provide sufficient numbers of appropriately skilled people in time, to maintain and enhance the quality of performance in terms of abilities, skills, knowledge and experience and to prepare for harmonised licensing.”

HUM02-ASP01 (Stakeholder Line of Action for ATSPs) states: 'Consider use of the First European ATCO selection test package which reflects future impacts and enables customised implementation and application'.

In the medium to long-term, communal efforts in initial and predictive test validation and in continuous improvement of quality, as is foreseen within the FEAST Work Package, will lead to:

- a generally higher quality of selection,
- harmonisation of selection in Europe,
- a more comparable level and profile of controllers.

1.2.1 Definition of Selection System

Under the considerations given, the STF agreed to the following definition of a selection system for ATCOs:

“In general selection constitutes a system of measures for identifying individuals whose personal and psychological qualities make them ideal candidates for controller training and job activities in the profession of ATC”.

The definition stresses the point that selection should aim at predicting both training and job performance of selected candidates and should take into account the requirements of training and of the controller’s job.

This document is based on the following assumptions:

- ATSP organisations are specific organisations requiring specific characteristics for effective performance.
- The identification of individual characteristics that are related to training success and operational efficiency are of ample importance in recruitment and selection.
- Different recruitment and selection practices attract and deliver different types of candidates with different suitability for the controller job into ATSP organisations.
- The future needs and requirements within the ATCO job will involve changes in the controller responsibility, more complexity, greater ambiguity and more rapid changes in the working environment, technical evolution

and changes in the working practices. These changes will reduce or at least change the meaning and suitability (usefulness) of traditional methods of recruitment and selection (for discussion of likely changes in the job and task of controllers see also Sub-chapter 13.3).

Current methods are challenged in particular because recruitment and selection still have to be made in the context of current existing systems, either in regard to training, the controller job or the work environment.

Within FEAST efforts have been made to capture recent job task requirements in a deductive Task Analysis (TA) approach. Studies undertaken within the 'Consequences of future Air Traffic Management (ATM) systems for air traffic controller Selection and Training (CAST)' Project gave indications as to the likely changes in future selection and training (see also CAST 1999a, b and c).

1.3 Task Force Activities

The STF included experts and professionals as representatives from nine ECAC States and from EUROCONTROL (Headquarters and Institute of Air Navigation Services [IANS]).

The Group adopted the "Guidelines for Selection Procedures and Tests for *Ab Initio* Trainee Controllers" (for consulting the revised version, please see EATMP, 2001b), produced by the former Selection Task Force I, as the general basis for its work. These guidelines outline the selection process and its components, and aim to give the principles of policy guidance for the definition of a selection system.

For implementation of appropriate methods and the evaluation of their effectiveness, detailed information and guidance is necessary and will be given in this document.

Professional best practices relating to implementation, administration and control such as:

- criteria for test evaluation,
- validations procedures,
- interviewing methods,
- administration of a selection system,
- characteristics of pre-selection procedures and
- maintenance of a database for the recruitment and selection stages

were adopted.

1.3.1 Task Force Work Programme

The STF agreed in its first meeting to the following work programme:

1. Define the general scope and content of work.
2. Define criteria for the evaluation of tools and methods.
3. Gather information by means of a survey questionnaire.
4. Evaluate available tools and methods.

1.3.2 Criteria for Evaluation of Selection Tools

In order to evaluate the potential effectiveness of available and emerging selection instruments, criteria were identified and defined. On the basis of these criteria, the STF designed a survey questionnaire, which aimed to gather relevant information on available tests, interviews and assessment centres.

A thorough analysis of those instruments which met the pre-conditions set by the group was done by independent experts (not by members of the STF) taking into consideration the weighted evaluation criteria seen by the group as of major importance and based on the information provided by the service providers.

A detailed description of the information on (a) tests in use and (b) tests and methods available for evaluation and/or use in ECAC States, as well as a description of the procedures applied in evaluating them and the result of the evaluation are available in EATCHIP (1997).

The Core Drafting Group for FEAST used the EATCHIP (1997) criteria for survey and for initial evaluation of tools and methods. Details on that evaluation will be included in a report on FEAST development, which is expected to be published in 2003.

2. TASK AND ABILITY ANALYSIS OF THE CONTROLLER JOB

The STF based its considerations on information available in regard to the tasks of controllers. The STF took note of various air traffic controller job and Task Analyses (TAs).

There is no single best way of performing a job analysis. Different approaches are linked to different theories and/or approaches and to what in a certain approach or theory is assumed to be important in terms of characteristics of persons in a certain job environment. Thus, they will lead to different sets of 'person descriptions'.

The FEAST, for example, adopted a **deductive** method, i.e. the use of job descriptor systems already existing (contrary to an **inductive** method, i.e. the development of a new tailored system). The deductive TA methodology revealed an integrated process model, which outlined the psychological factors that are significant for becoming an ATCO and are relevant for selection. Further details will be published in the FEAST manual.

2.1 Guiding Principles

The objective in selection is to recruit, select and employ those candidates who have shown their potential for the job as an ATCO.

There should ideally be a close connection between the requirements in selection, in training and in the tasks of an ATCO (see Figure 1).

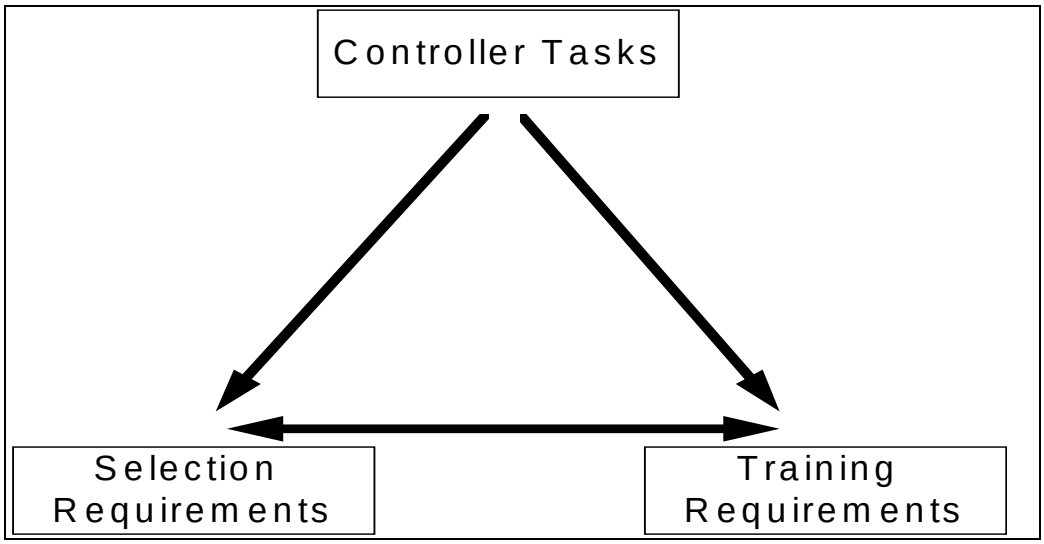

Figure 1: Links between selection, training and controller task requirements

Note: Figure 1 does not suggest that selection criteria can be directly derived from training and from the controller task. In fact, this would be the last step in the whole process. The first thing to happen is to identify the task requirements by an appropriate TA. The next step should be to transform task requirements into training schemes or scenarios. In a third step selection tools can be designed, developed or identified using both the information from TA and from training.

What is done in selection should ideally reflect the requirements in training and in the tasks of ATCOs; training requirements should at the same time ideally reflect the task requirements on the job.

It is obvious that this ideal situation does not exist. In fact, the current situation is marked by the non-existence of a valid and universally applicable TA of controller functions. However, the STF adopted the following general principle as the basis for a best practice guideline:

- a first step in designing or developing a test battery is to carry out a TA from which would emerge not only the basic knowledge and skills required for a specific type of operation, but also the determination of the ways in which these skills are combined.

The aim of a TA should be to understand how controllers perform their job. This information would serve to obtain better measures of performance in training and on-the-job performance as selection criteria.

2.2 Task and Job Analysis

For clarification, definitions of terms used are listed below (see also Annex "Glossary of Terms"):

- **Job Analysis**: The systematic collection of data describing the tasks that comprise a job and the knowledge, skills, abilities and other characteristics that enable an individual to carry out those tasks (Landy, 1993).

- **Task Analysis**: A group of analytical procedures directed at describing work activities in detail. A TA aims to provide a comprehensive breakdown of the content of tasks.

For further discussion on different approaches of TA, please refer to the Technical Supplement (TS), Chapter 2.

Guideline 1: Selection Tests Linked to Controller Task

A Task Analysis should precede the design or development of a selection instrument or the composition of a test battery using available tests.

Guideline 2: Task and Job Analysis

As far as selection is concerned a Task and Job Analysis should be conducted aimed at identifying the cognitive and social abilities, knowledge, skills and responsibilities necessary in the tasks and the job of air traffic controllers.

However, implementing Guideline 1 into any selection system could be a troublesome task requiring effort and skill. The STF recognises that to conduct properly a TA is a complicated, time consuming and expensive activity. It is almost impossible to carry out a new TA every time there is a small change in the job. However, for future selection needs, whenever serious doubt arises that the current TA

- is out of date or
- has not yet been proven to be transferable (i.e. in a cross cultural sense) or
- is not considered to have been comprehensive enough (i.e. does not cover all responsibilities of a controller) or
- does not cover future development and changes in the controller task,

the STF would encourage and highly recommend selection authorities to carry out a new TA within their respective ATC environment.

This recommendation is in line with what was mentioned in Sub-chapter 6.1 of EATMP (2001b).

2.3 Differences Between Controllers' Jobs

Clear criteria for the existing different jobs in ATC are needed for the purposes of selection. All controllers issue instructions to ensure aircraft separation; all communicate with pilots and coordinate activities with other controllers. However, their job functions differ, e.g. the aircraft they control are moving at different speeds, they control different numbers of aircraft, or they may or may not use radar as a tool.

These differences, if identified, could be taken into account in an advanced selection and training system where early assignments to different jobs are available. Because of shortcomings in the MP process current selection is in many instances not geared towards selecting controllers for specific jobs.

Guideline 3: Selection and Training Geared towards Job Differences

If it is intended that selection, training and Manpower Planning policies take into account differences in job functions, then task analysis tools need to be developed to fully identify these differences.

2.4 Future Job

TA can be applied either to the current work or future work of controllers. The requirements of the future job, if identified, could also be taken into account in an advanced selection and training system. Current selection is almost completely geared towards selecting controllers for existing technical systems and training schemes due to the fact that future requirements in the jobs are not yet completely identified. However, first attempts have been made to investigate abilities / skills changes stemming from technological changes, e.g. controller-pilot datalink technology (see CAST, 1999a, b and c). More efforts are required to identify likely changes in the required profile, the roles and responsibilities and related skills of controllers in the short to medium term.

A TA for future work employs deductive, hypothesised methods based on objectives, facilities, rules, procedures, instructions, envisaged skills, experience and knowledge. (See also Chapter 13, “Future Trends and Steps”.)

Guideline 4: Selection and Training Geared towards Future Jobs

Efforts should be made in Manpower Planning, selection policy and training to take into account future requirements in the controllers job as identified by Task Analysis or other means.

3. TRAINING AND ASSESSMENT ISSUES

This Chapter outlines training and assessment issues linked with recruitment and selection.

The STF based its considerations on some information which was available on current procedures in training and assessment of *ab initio* trainees. The STF recommended to note the work done within EATCHIP in the Human Resources Training Sub-Domain with regard to

- training content and training objectives (ST05, Specification of Training);
- training methods and training tools (ST07, Application of Training Methods and Provision of Training Tools; ST12, Development of Air Traffic Services (ATS) Illustrative Courses and/or Related Courseware).

TSP-002-006 of the EATMP HRS Programme outlines new training technologies and methods. On-the-Job Training Instructors (OJTIs) training and training assessments are stressed. The use of advanced training methodologies will enable shorter training, part-task training and will include Pre-On-the-Job Training (Pre-OJT) (simulation). OJT (Unit Training) is the most crucial training stage. For further details please consult the Training Sub-Programme (TSP) deliverables.

3.1 Guiding Principles

3.1.1 Links Between Selection and Training

Training and assessment issues linked to selection are important to be considered in this framework because candidates do not take up the job of a controller immediately after selection. There is a long training period between selection and actual performance on the job.

Selection of *ab initio* trainee controllers has therefore the ultimate aim to select candidates who can be trained in a given system and have the potential to do the job.

A dilemma would exist for selectors if training and job requirements were in contradiction or inconsistent. A clear understanding of the performance criteria in training and in the job is an inevitable precondition for any selection system to assure that both goals can be achieved.

A similar precondition applies to training: training has to build on the basic Knowledge, Abilities, Skills and Other characteristics (KASOs) identified in the recruitment and selection process, by adding the knowledge, skills and proficiency necessary to meet the demands of the controller job.

A dilemma would exist for trainers if those entering training lacked the necessary basic competencies to build on. A clear understanding of the candidate profile is thus an inevitable precondition for any training system.

Thus, it has to be stressed that there is no room for a complete independent existence of both systems as they aim at dependent goals:

- selection aims at identifying the potential of candidates necessary for later success in training and in the job (selecting for success);
- training aims at making the full use and build-up of the potential of candidates which is required to be a successful controller (training for success).

The performance in each stage must predict the performance at the next stage.

Guideline 5: Close Links Between Training and Selection

Close interdependencies exist between goals in selection and in training. Close links should be established between both domains ensuring a high degree of coordination and cooperation thus leading towards success.

3.1.2 Basic Assumptions

This Chapter is based on the following assumptions:

- a close link exists between selection and training and therefore close coordination between selection and training activities is necessary;
- changes in selection policy or methods influence training in many respects and vice versa;
- the effectiveness and efficiency of the selection system depend to a high degree on the effectiveness and efficiency of the training system and vice versa and on feedback from training.

3.1.3 Training Objectives and Activities Linked to Selection

ATC training can be defined as follows:

"On the basis of trainees' basic abilities and skills identified in the recruitment and selection process, ATC training is a set of planned activities on the part of an ATSP organisation to give or build-up the job knowledge, skills and proficiency and to modify the attitudes and the social behaviour consistent with the requirements of the job and the goals of the organisation" (amended after Landy, 1989).

The following major activities in training are related to selection; some of them will be tackled in more detail in Sub-chapter 3.2 on policy recommendations:

- identifying and defining performance criteria in training aligned with job requirements and based on the profile of selected candidates;
- choosing appropriate training content;
- choosing appropriate and effective training methods, techniques and media;
- choosing appropriate length of training;
- developing and applying appropriate, reliable and valid methods for ongoing performance assessment of individual trainees in all training phases and including Pre-OJT and OJT;
- following up and gathering data on trainees throughout the whole training for later evaluation and validation of: training methods and media, performance assessment methods, recruitment and selection methods and tests;
- evaluating training progress and performance in Pre-OJT and OJT at regular intervals.

3.2 Policy Recommendations

3.2.1 Performance Criteria in Training

The STF felt that one of the critical issues in selection is the lack of clear, measurable, stable and valid performance criteria in both controller training and on the job which sufficiently represent the complexity of the controller job.

It is necessary, however, to stress the following point: the efficiency and effectiveness of selection tools depend to a great extent on the availability of such criteria - both in training and the job - against which tests can be validated.

The practice, which is still in use in some States, is to simply state that the (overall) performance criterion in training is to pass training. Detailed assessment scores are available from the Initial Training phase but rarely from Unit Training in most States. Other scores or scales, even if they exist, are not always available for validation purposes. The pass / fail criterion is therefore in most instances the only criterion available for validation of selection procedures and tests (see Chapter 4, “Validation Issues”, for detailed considerations, and Chapter 5, “The Scope of Multiple Regression Methods”, in the TS).

The STF felt that efforts should be made to overcome this unsatisfactory situation and to establish performance criteria and the (minimum) performance level to be achieved in different training phases:

- in Initial Training: learning the theory, acquisition of knowledge;
- in Unit Training / On-the-Job Training (OJT): application of knowledge acquired, acquisition of practical skills.

This performance criteria in training should

- reflect the complexity of the job at different complexity levels addressed in different training phases;
- be stable (reliable) and valid (e.g. relevant for success in training and for operational proficiency);
- allow for the development and application of reliable and standardised methods in the measurement and observations in performance assessment of trainees.

Performance criteria need to be identified and defined in close cooperation between training, operational and selection units. A performance criterion could be, for example, an average of scores in simulation. See also Sub-chapter 4.3.1).

The CDG invested efforts in the development and validation of criteria for trainee performance in different training phases, including OJT.

Guideline 6: Performance Criteria in Training

Relevant, stable, valid and measurable performance criteria in training should be identified and agreed with operational and selection units which can be used to arrive at an appropriate design of training and assessment procedures and which can be used in the validation of recruitment, selection and training methods and tools.

3.2.2 Length of Training

Choosing an appropriate length of a training course is a complex task. Considerations will be given to training content and complexity, training methodology used and financial / administrative issues against the average capabilities of trainees in terms of pre-entry requirements, i.e. education, age, pre-training experience and level / profile selected.

For economic reasons, training will be designed so that the final fitting of a candidate for full employment is achieved in the shortest possible time.

From a purely statistical point of view, there is strong evidence that any extension of a training course will mean that the correlation between

pre-training selection and performance on the job is decreased. The converse is not true, shortening a training course will not necessarily increase this correlation.

3.2.3 Training Content and Method

The further the training curriculum is from the actual situation 'on site', the less effective the training is in predicting success on the job.

There is some evidence, to review the strict distinction (see above) between Initial and Unit Training in the light of training failures and the difficulties some of the trainees have in progressing from learned theory to practical application of the knowledge acquired.

Some trainees may be successful in the classroom but may lack the capacity to automate certain perceptual and motor aspects of the controller's task and may fail at simulator training which requires fast and efficient output.

Moreover, although those highly skilled in this respect may successfully transfer their skill to OJT, they may nevertheless fail overall because of difficulty

- in switching rapidly from a vigilance mode during periods of low activity to a fast output mode during periods of heavy activity and/or
- in maintaining concentration in the vigilance environment (Heron et al., 1995) and/or
- in coping with emotional problems.

This points to the fact that trainees have an additional need of the specific knowledge, skills and abilities to be able to control their workload as well (Haglund et al., 1996) and concerns therefore the more general question of training objectives and training content[1].

The STF felt that special consideration should be given to the content of the institutional training in regard to the following:

- TA as the basis for training design;
- Initial Training should give students a systems view (e.g. to form mental models about ATC as a complex system);
- the early introduction of practical exercises (e.g. Computer-Based Training [CBT] or simulation) to practise ATC from the beginning of training and at different levels of system complexity in subsequent training phases;
- teamwork aspects.

[1] These issues are developed in the "Guidelines for Common Core Content and Training Objectives for Air Traffic Controllers Training (Phase I: Revised)" [EATMP, 2001c] & "(Phase II)" [EATMP, 2000c).

It should now be noted that Training Plans devised for "Rating Training" for *ab initio* controller now include considerable use of CBT items and an extensive use of a Radar Skills package (Part-task Training), both of which will operate without the use of a large scale simulator. It is intended in this fashion to use analytical exercise to highlight specific objectives, and to amalgamate the outcome from this analytical element into global type exercises where the "parts" can be fused into the "whole".

3.2.4 Performance Assessment Methods

There is also evidence for the fact that some of the drop-out rate is due to inappropriate assessment procedures in use in training systems.

One of the problems in performance assessment of trainees is the fact that appraisal and assessment is often done in psychological or psychiatric terms pointing towards trainee personality traits or other attributes instead of in terms of training progress (Haglund et al., 1996).

Assessment should

- be done on an ongoing basis and should ensure that trainees meet earlier objectives before proceeding with the course;
- reflect the complexity of the training task or exercise;
- ensure reliable procedures (e.g. by increasing reliability of instructor or coach assessments or judgements, by avoiding any lack of uniformity of standards and competence among trainers, instructors and coaches in the administration of assessment tools);
- allow for the application of standardised methods for assessment of trainee competence and behaviour (e.g. observation scales that can be used in all phases of training);
- assure that the skills and qualifications of trainers, instructors or coaches are clearly identified and defined and that appropriate training in applying performance assessment tools is given;
- ensure appropriate feedback to trainees;
- clearly indicate to both, trainees and instructors, areas for improvement.

Guideline 7: Assessment Procedures

Objective, reliable and standardised methods for performance assessments in courses or methods for performance measurement in simulation should be put in place and should be used on an ongoing basis by appropriately trained instructors or coaches.

3.2.5 Trainee Follow-up

In order to follow up on the performance and success of all trainees in all phases of Initial and Unit Training, it is necessary to gather data in as many training variables as possible.

Information about the training and examination scheme that has an influence and/or is related to training success should be gathered for each individual trainee as, for example:

- scores of examinations during training;
- number of hours in Initial Training;
- number of hours in Unit Training / simulation;
- number of actual hours in OJT at the appropriate level of complexity;
- whether the trainee had resit and in which training phase(s);
- whether there was re-coursing or allocation to different sites / jobs within ATS;
- reasons for training failure, for resits, re-coursing or relocation.

This data should form part of an integrated recruitment, selection and training database. The structure and content of this database is outlined in Chapter 12, "Recruitment, Selection and Training Database".

Guideline 8: Procedures in Trainee Follow-up

Appropriate amounts of relevant data should be gathered on an ongoing basis and throughout the whole training to follow up on the progress of trainees. This data should form part of an integrated recruitment, selection and training database.

3.2.6 Evaluation and Validation Procedures

The procedures to be followed in training evaluation and validation should follow similar lines as outlined in Chapter 4, "Validation Issues", and in Chapter 12, "Recruitment, Selection and Training Database".

Page intentionally left blank

4. VALIDATION ISSUES

4.1 Introduction

A selection procedure should have clear objectives, consistent with legal, professional and ethical responsibilities. Validation determines how well these objectives have been achieved.

Validity refers to the appropriateness, meaningfulness and usefulness of the specific inferences made from test scores (Committee to Develop Standards for Educational and Psychological Testing [CDSEPT], 1999).

Validation is the process of accumulating evidence to support such inferences. The inferences regarding specific uses of a test are validated, not the test itself (CDSEPT, 1999).

Since evidence may be accumulated in many ways, there are different approaches of measuring different types of validity. Chapter 4 of the TS presents the main validity strategies.

Validation means the confirmation or establishment of something. Validity in a testing context is the extent to which predictor samples of behaviour effectively overlap with performance domains. A predictor is a source of help to selection decision-making. It can include, for example, biographical data, standardised ability tests, ratings based on interview, structured behaviour samples, etc.

A test is, therefore, valid if it measures what it is intended to measure. This is not as straightforward or technical as it might first appear. In fact, generating validity evidence is a complex process. A test can have few validity coefficients and can be more valid under certain circumstances than others. It has to be demonstrated empirically that a test in use is in fact valid.

Guideline 9: Validity

Inferences made from test scores of *ab initio* trainee controllers require evidence of validity. A well-executed, relevant, large-sample validation study, where methodological and statistical artefacts are controlled is a prerequisite to any selection test. A periodic and carefully designed validation study is a measure to ensure the quality of selection.

A test cannot be valid if it is not first reliable. In particular, scoring variations, which affect reliability, may occur in work samples. However, a reliable test is not necessarily valid. Reliability is necessary but not sufficient for validity. A valid test should also have, for example, the power to discriminate between candidates.

4.2 The Utility of Validation

Routine validation studies, not to mention complex validation studies, are quite expensive. However, high selection validity translates into considerable financial savings. The results of a failure to employ a valid test in selection because its validity was not verified could be even more expensive. Usually the cost of testing (including validation and test maintenance) is insignificant relative to the overall selection utility (Anastasi, 1982). The cost of paper-and-pencil tests is fairly low in comparison to the costs of unsuccessful candidates in training or in OJT.

Random selection, though rarely if ever practised, provides a baseline for comparison of utility gains and losses. Even a 'low' significant validity coefficient of .25 (or .20) can have a practical value, in generating substantial savings as a result of selection.

Psychological testing provides a tool for more equitable access to employment. Although not all tests are well developed, or beneficial, proper use of well-validated tests provides a better basis for making decisions about individuals than would otherwise be available.

Value judgements are usually involved in selection decisions. Whether the most important error to avoid is a false positive (selecting a person who subsequently fails) or a false negative error (not selecting a person who would have succeeded) is a value judgement which depends on the cost assigned to each error. See also Chapter 4 of the TS (1998a), the guidelines document "Characteristics of Recruitment and Pre-selection of *Ab Initio* Trainee Controllers (Revised)" (EATMP, 2001d), and the report "A Framework for Applying Cost-Benefit Considerations in the Recruitment and Selection Process for *Ab Initio* Trainee Controllers" (EATCHIP, 1998b).

4.3 Methodological Problems

The methodological problems mentioned below are not strictly theoretical; they might pose legal threats to the use of some selection tools as predictors of work behaviour. Some of the factors which obscure test validity and may tend to reduce the efficiency of a validation study in uncovering predictor validity are criterion measures, sample size, criterion unreliability, restriction of range, two-tailed test, intercorrelations, adverse impact and sub-scores. Sub-chapter 4.3.1 describes the "Criterion Measures" problem while Sub-chapter 4.4 of the TS addresses the other issues in detail.

4.3.1 Criterion Measures

Criterion measures and the rationale for choosing them as relevant are preferably described explicitly in a validation report since they provide significant information.

The following factors contribute to the error in measuring performance and thus hamper the validation of selection tools.

The first question in any validation is what criteria should be predicted. A decision has to be made whether **training** results or **job** performance should be predicted by selection methods. As was mentioned in Chapter 3, **criteria** for job performance are still not clearly identified and thus valid **measures** of job performance are not available.

The type of the **criterion measure** used is an important issue too. If a global training (i.e. pass / fail) or job **performance criteria** is used it is obvious that tests measuring a restricted range of abilities cannot be highly related to the criteria. It could be argued therefore that several criteria should be used in validation.

Besides sufficiently high reliability and validity of the criteria it is also important that the criterion measure differentiates between able and less able trainees or job incumbents, and that it still has a sufficiently high variance. This is obviously not the case with a **pass / fail criteria**. Performance criteria allowing for variance and measured on a continuous scale are appropriate.

The problem of defining dependent variables beyond the pass / fail judgements of the instructors and managers at the ATC Academy is of major concern. In recent years there have been developments in some European States to assess students' performance based on training objectives to be achieved in a process of **continuous assessment** - improvement instead of static judgements.

This Sub-chapter outlines types of **dependent variables** and their corresponding statistical calculations. See also Chapter 5 of the TS on multiple regression.

On an **ordinal scale** one could, for example, develop a rank order on each phase of the training, and a rank order among all students being trained at the site. This kind of variable can be submitted only to ordinary statistical processing. The use of ordinal measures as though they were interval or ratio measures, under the assumption of interval equality (equal intervals on the scale) can err on the side of interpreting data and the relations inferred from data. This danger is grave enough to recommend avoidance of this kind of assumption.

A preferred option would be to elaborate dependent variables which could be used on an **interval scale**. These could be, for example, some behavioural measurements in simulator exercises, such as number of uncomfortable steep climbs, number of aircraft on too high altitude and others. During OJT the number of days / hours spent in training (after deducting factors beyond the trainee's influence such as operational requirements, weather conditions, etc.)

Guideline 10: Validation Criteria - Use of the Pass / Fail Criterion

In validation the commonly used pass / fail criterion might be sufficient for justifying a failure case, but not detailed enough to describe specific strengths and weaknesses in training performance for validation purposes. Although most validity studies of Air Traffic Control training use a success / failure criterion, training success as a criterion for validation should refer as much as possible to actual grades achieved in class or simulators.

In the early validation of ATC test studies the criteria were usually lecture grades and instructors ratings (Brokaw, 1984). Later studies used supervisor ratings, course grade composites, as well as attrition, career progression and disciplinary actions as criteria. These studies used the supervisor ratings as the 'real' measure of job performance and the other criteria were evaluated on how effectively they predicted the supervisor ratings (Manning, 1991).

At the ATC training academy students usually learn some basic aviation and ATC concepts. They are tested on their ability to apply these concepts in high or low fidelity simulated problems. Records of training institutes obtain, in many instances, scores on training tests as well as number of errors made in different categories on each simulation exercise. These scores are commonly in use as criteria for performance during the training period. Since each stage in this multi-stage system should predict the next stage (see Sub-chapters 10.2 and 11.1) it would be interesting to have criteria of performance on the job itself and not only under 'laboratory' conditions at the ATC academy. Furthermore, the training scores criteria are actually measuring performance on the learning curve and not asymptotic performance (Manning, 1991).

The technical assessments of former ATCOs - who are in many cases the instructors - are subjective ratings. Despite serious efforts to design standardised instructor rating procedures, it is almost impossible to ensure that observations of performance occurring under varying conditions are conducted in a standardised way (Manning, 1991).

However, the training philosophy in Europe is perhaps quite different. Within most European States a change in this respect has taken place towards more objective-oriented training, continuous assessment, identifying the less developed students and assessing the reasons for failures and not only the failures themselves.

Guideline 11: Training Scores as Criteria

When training scores are used as a criterion measure, it is necessary to ensure their job relevance. The development of supervisors rating scales should ordinarily be guided by job analysis. Usually, it may also be necessary to train raters in using the rating scales in a standardised way (Society for Industrial and Organizational Psychology, Inc. [SIOP], 1987).

Information is gathered about *ab initio* trainee controllers regarding different phases of training they have completed, including dates, number of OJT hours required and grades or rating of their performance on pre-defined criteria. The FEAST development included an identification of key criteria that are relevant for performance in the ATC job and are thus relevant for selecting new recruits for the job.

Sector-dependent Recordings

The recorded information may be 'sector-dependent', that is to say relevant to the sector on which the trainees were trained, even if not necessarily the one on which they were certified. The number and types of aircraft, and their configurations in the airspace, differ from one sector to another. On-the-job trainees are not necessarily assigned to sectors in the order of complexity (from the easy one to the more difficult). They very often have a different number of instructors even in the same sector and on the same day. They are trained under Instrument Flight Rules (IFR) or Visual Flight Rules (VFR) conditions. The frequency of OJT is different between centres not only due to operational requirements, but also due to training tradition, the importance accorded to training and the organisation's philosophy.

Sometimes trainees pass the non-radar scenarios but fail the subsequent radar training or vice versa.

Radar training may occur after non-radar training at terminal facilities, whereas non-radar training can occur at the same time as radar training in the en route training. All these variants have direct implications on the validation of selection tools.

ATC is considered to be almost an art or craft rather than a standardised job (Manning, 1991). ATCOs have the freedom to utilise different individual techniques to control aircraft as long as their techniques safely and efficiently separate aircraft and conform to the generally defined operating procedures.

Although both **en route** and **terminal controllers** formulate clearances to ensure aircraft separation, both communicate with pilots and coordinate activities with other controllers, they can differ in the different speeds and stages of converging on an airport or the aircraft they control, the numbers of aircraft they control and the use / non-use of radar.

Where possible, objective criteria should be preferred for validation purposes. For example, when Initial Training and OJT are both evaluated based on **subjective** ratings, there could still be a high correlation between the results of both types of training, but reflecting only the degree to which trainees have got along with their instructors both in the institute and in the field. Although this correlation could also predict to a high extent the degree to which these trainees will also get along with their future colleagues or supervisors, it is difficult to use it as an **objective** criterion for validation purposes.

4.3.2 Other Criterion-related Obstacles

Validation studies face additional criterion-related obstacles when:

- the factor structure of job performance is different between jobs (e.g. two ATC positions in the same sector, e.g. one may be less busy than another; another may be seasonally affected);
- validity is specific to situations (e.g. resolution of potential en route conflicts, but not of the terminal manoeuvring area);
- the supervisor or the evaluator of performance cannot detect subtle differences between two controllers (within FEAST an assessment tool was developed to address the obstacles described in the last two bullets);
- validities tend to decrease over time.

4.4 Composing a Test Battery

A test battery is composed of different tests and other elements. Introducing and adapting a test battery developed in another country or composing a test battery using tests from different sources needs to be done with care.

4.4.1 Use of a Complete Test Battery

If a complete test battery is taken with proven validity and reliability, the question is still whether the reported figures can be expected in a new setting. Consideration has to be given to the following (and probably to further) factors:

- language environment (e.g. if test has to be translated into another language);
- recruiting methods and/or sources;
- educational settings;
- applicant population due to age, gender or nationality;
- training scheme and training methods in use;
- performance criteria to be predicted;
- task to be performed, working environment;
- suggested cut-off scores to be used.

4.4.2 Use of Single Tests from Different Sources

The task is more complicated if tests from different sources are selected and made up into a new test battery. The following uncertainties exist in addition to the factors mentioned above:

- the interrelations (correlations) between tests are unknown; the amount of common variance and the variance explained by each test has to be established;
- the learning effect from one test to another;
- the composite validity of the resulting test battery has to be established as well.

It is obvious that in both cases the test battery / single test should only be introduced after thorough consideration and proof of the mentioned factors.

Guideline 12: Composing a Test Battery

A test battery should be put together by psychologists well experienced in test development and test application.

Details of possible ways of introducing tests are given in Chapter 3 of the TS.

4.4.3 Use of Tests for Different Job Levels

Where a selection procedure is also designed for a higher level job (e.g. supervisor) than that for which candidates are initially selected (e.g. *ab initio* trainee controller) and the majority will not advance to the supervisor level, it may still be acceptable to assess candidates for that job, if the validity study includes criteria that reflect performance at the higher level along with criteria for the entry level (SIOP, 1987).

4.4.4 Content of a Validation Research Report

In summary, an ideal validation research report will state clearly, accurately, in detail and in a language likely to be interpreted correctly, identifying information on the

- author,
- date of the study,
- improvements which the new selection procedure are expected to achieve,
- results of the job analysis, if conducted,

- description of criterion measures, their collection procedures, their relevance, reliability and the rationale behind their use,
- characteristics of the research sample (e.g. representativeness of population, sampling biases and restriction of range),
- names and editions of selection procedures,
- description of relationship to the job performance,
- references to related studies,
- complete statistical data (including uncorrected values),
- sample size,
- measures of central tendency, variability and intercorrelations,
- any transformation of raw scores,
- recommendations (e.g. cut-off scores),
- information material developed for users and test takers,
- warnings against common misuses or incorrect interpretations,
- any specially required qualifications of administrators or assessors,
- instructions for standardised administration, scoring, scaling and interpretation, and on
- the interval of time that passed in collecting the criterion data (SIOP, 1987).

Computer-based (and Web-based) test interpretations are validated using the same guidelines mentioned above.

4.5 Validation Strategies

Ideally from a theoretical point of view, in order to analyse the validity of the test's predictions, it is necessary to take candidates with the total variance of the applicant population and then monitor their performance in the criterion variables. It is also possible to determine the validity of a test's predictions by taking random individuals who fall below the cut-off value in the tests. However, in practice it is costly and may affect safety.

Validity problems must also be analysed against the variables with which test values are correlated: Time changes the relationship between predictors and criterion. There might be a close relationship between certain predictors and the results in the theoretical test during training, but the relationship decreases

as one concentrates more on proficiency variables (e.g. in OJT). (The reverse might also be true.)

It is therefore best to measure trainees during the whole training (and if possible after OJT) in as many variables as possible (see Sub-chapter 12.3 on database structure).

Another method of controlling and improving the validity of a test is to compare test results for recently examined (checked out) controllers with assessments from the training course. It should be borne in mind that due to influences during training there could be differences in personality and skills between applicants. It would be useful to compare different groups (e.g. *ab initio* at the beginning of the training and others who recently checked out).

Since the aim of testing and selection is both to admit applicants who can be trained to become a controller and to admit persons who are suitable for controller work, i.e. who can control traffic in a safe, expeditious and economic way, the validity of the test battery will also be dependent on differences between trained controllers. If the effect of controller training, technical systems and the methods used is that differences between controllers are minimal as regards job proficiency, the validity of the relationship between test results and subsequent job proficiency will also be limited owing to the minimal variance in the criteria.

Details in regard to possible ways of test validation and statistical procedures to be applied are given in Chapter 4 of the TS.

4.5.1 Comparison and Combination of Predictors

For most jobs there is no known test which is as good as **cognitive tests** in terms of validity. Numerous studies have shown that cognitive tests are the most valid predictors of most jobs (Hunter and Hunter, (1984); Schmidt and Hunter, (1998)).

Sub-chapter 4.6 of the TS provides more details on comparison of predictors.

Some of the additional tools to cognitive or perceptual ability tests (e.g. predictors which measure social skills or personality traits) that are relevant to the ATC job can increase the total validity of the test battery.

The contributions of various components to the variance of the composite score of the test battery are unlikely to be the same. Irrelevant components should be minimised. Parts that do not contribute to the total variance should ordinarily be eliminated. Effective weights depend on the standard deviations and the intercorrelations between the components. Part of the validation will include the judgement involved in the weighting decision (SIOP, 1987).

The validity of an inappropriate combination of predictors can be lower than the validity of the best single predictor.

Guideline 13: Combination of Predictors

Predictors should not be weighted equally, but rather combined in accordance with their actual validities.

Alternative predictors can be compared by computing the utility of the selection method. One can compare, in currency terms, the production of those hired by a specific selection procedure with the production of the same number of employees hired randomly. This kind of comparison depends on the **selection ratio** and the standard deviation (reflecting individual differences).

4.5.2 Selection Strategies

One could select solely on the basis of ability without regard to sub-group identification (e.g. males) under the assumption that ability predicts probable job performance. This strategy will maximise the utility. One could also select on the basis of ability within each sub-group separately, i.e. selecting in exact quotas in order to satisfy **'affirmative action'** goals. A third strategy would be to use ability tests with a very low cut-off score (Hunter and Hunter, 1984).

The use of low cut-off scores does not equalise hiring rates and also negates most of the gain derived from the predictive power of ability tests by hiring from the top down. The utility of low cut-off scores is also inferior to quotas (Hunter and Hunter, 1984).

This may be a significant issue in the US, but for the time being to a lesser degree in the European context.

Random selection from among those scored above a cut-off score, typically reduces the validity and utility of the selection procedure, when the true relationship is linear (Schmidt, Mack and Hunter, 1984). For further details on cut-off scores see also Sub-chapter 10.1.

Guideline 14: Top - Down Selection

In normal circumstances, selecting from the top scores down, if there is an appropriate amount of variance in the predictor, is almost always the most beneficial procedure (SIOP, 1987). However, this approach is appropriate only in cases of sufficient numbers of suitable candidates.

4.6 Is a Validation Study Feasible?

In designing a validation study one should not only identify the nature of appropriate samples; the feasibility of specific statistical controls or corrections and the number of cases, but also there are some specific questions to be asked before validating a test for *ab initio* trainee controllers. Some of these questions are presented in Sub-chapter 4.7 of the TS.

4.7 Which Kind of Validity to Use?

An examination of the professional literature indicates that no consensus exists regarding superiority of one form of validity or another, judgemental or empirical. Sub-chapter 4.8 in the TS presents a critical discussion to the reader and not a judgement on the superiority.

Guideline 15: Sets of Validity

As long as the evidence involved is based on sound professionally accepted standards for determining the validity, in interpreting test validity one should take into account a set of validities and not just one figure, or one coefficient or one type of validity.

Guideline 16: Combination of Validity Evidence and Choice of Strategy

A rationale for a specific combination of few kinds of validity evidence should be specified in a validity study while taking into account the questions of the study and the context of the evidences. The rationale should also be mentioned where a choice was made between either a predictive or a concurrent validation strategy.

It is important to stress that, although test developers should supply the required information on validation, the ultimate responsibility for appropriate test use lies with the user who requires the test results for decision-making purposes. Responsibility cannot be abrogated by purchasing services from an outside provider.

A validation study is not carried out for its own sake. The results of the study will serve to assess the practical value of a selection tool or procedure.

Page intentionally left blank

5. FRAMEWORK FOR IMPLEMENTING A RECRUITMENT AND SELECTION PROCESS

5.1 Introduction

Recruitment and selection of personnel for the job as controller is an essential part of the wider personnel and Human Resource Management (HRM) activities of ATSP organisations. In addition to this, controller training is vital for the success of ATSP organisations and should be an integral part of the HRM activities.

It is strongly felt that these activities should be closely and consistently linked with the business of an ATSP organisation to be effective and that HRM activities should be used to manage the Human Resources (HR) an organisation is trying to obtain.

This Chapter highlights the importance of mission, objectives and goals in recruitment and selection, the competencies, roles and responsibilities of staff in the process and outlines important policies that need to be considered and adopted.

The STF wants to stress the point that the policies and the procedures that are applied in recruitment and selection should be agreed within the organisation and should be documented in a manual, in such a way that it gives guidance and provides support to staff involved in the recruitment and selection process.

5.2 Mission, Objectives and Goals in Recruitment and Selection

A **mission** statement should encompass all goals, activities and plans in recruitment and selection putting them into perspective with the general aims and purposes of the organisation.

Objectives are qualitative statements that define the specific needs to be met in recruitment and selection.

Goals are quantitative statements that give targets to be achieved in recruitment and selection.

5.2.1 Mission

A high-level mission statement on recruitment and selection has to be agreed within the ATSP organisation in close collaboration between the units for MP, training and operations.

It should state:

- how recruitment and selection are linked to the strategic goals of the organisation;
- the general aims and the purpose of recruiting and selecting ATC personnel;
- the links to other internal and external service units in MP, training, operations and licensing;
- what service should be provided in the selection process;
- the quality system applied in the process of service delivery.

The mission statement must be kept regularly updated to ensure the maintenance of service quality.

Due regard should be given to any foreseeable changes in the organisation (e.g. corporatisation, privatisation, reorganisation, management), changes in the services offered by the ATSP organisation, technical evolutions, etc., that have an impact on the profile of the people to be recruited and selected.

5.2.2 Objectives and Goals

Recruitment

The general objective of recruitment is to ensure that the right kind of candidate is targeted, attracted to the controller job and the organisation, and finally applies. More specific objectives will describe in qualitative terms

- what population should be targeted (e.g. profile and characteristics of the target population);
- what is seen as a successful means to attract people for the job (e.g. information content and media);
- what is seen as a meaningful application (e.g. content of application).

Consideration should be given to influencing, existing factors and more specific objectives, which should be derived from the principles and policy recommendations outlined in Chapter 6, “Recruitment Issues”.

The specific goals to be achieved in recruitment will vary from State to State and cannot be given here. (See further details in EATMP [2001d]).

Selection

The general objective of selection is to ensure a choice of candidates who meet the requirements for the controller job, including training.

Specific objectives in selection will describe in qualitative terms:

- how the knowledge, skills and abilities of applicants required in training and in the job (e.g. job analysis, Critical Incident Technique [CIT]) will be established;

- what are the basic and advanced requirements applicants must fulfil in order to be pre-selected for further testing (e.g. age, education, medical);

- what is seen to be the purpose and means in further assessing the abilities, skills and knowledge of applicants (e.g. psychometric tests, interviews);

- what qualities and to what level (e.g. high, medium or low) applicants should have in order to be selected as candidates for *ab initio* training.

The specific goals to be achieved in the selection process will also vary from State to State. This will most probably concern the

- selection ratio (see Annex "Glossary of Terms") and target figures in pre-selection and selection to be fulfilled;

- timescales and deadlines to be achieved in pre-selection and selection (e.g. maximum time allowed between receipt of application form and intake in training);

- financial targets (e.g. average costs per selected candidate);

- success targets (e.g. ratio of successful to selected candidates in training).

Some important additional goals should concern the level of quality of the selection service, the quality of assessment instruments (e.g. validity, reliability), the necessary level of staff training required and their competencies.

Objectives and goals need to be monitored and regularly updated and/or adapted to the situation as required.

Further details and recommendations are given in the chapters of this document dealing with validity (Chapter 4) and training requirements (Chapter 3).

Guideline 17: Mission, Objectives and Goals in Recruitment and Selection

The mission, objectives and goals of recruitment and selection need to be defined and agreed internally in close collaboration with the units responsible for MP, training and licensing.

Regular updates and/or adaptations of mission, objectives and goals should be done and should ensure that a close link to the business needs of the organisation, intended changes and evolutions exists at any time.

5.3 Policies in Recruitment and Selection Management

5.3.1 Recruitment and Selection Policy and Plan

A recruitment and selection plan is an essential part of the MP process and must be agreed on a collaborative basis with the training and MP units. The selection plan must allow for the intake of sufficient numbers of candidates within a certain time period.

A balanced recruitment and selection will ensure that the intake of candidates for training will be evenly distributed over a number of years thereby avoiding a large age pyramid.

The recruitment and selection plan should avoid the inefficiencies of a stop / start approach to hiring (see Sub-chapter 7.1 of the Report “ATS Manpower Planning in Practice: Introduction to a Qualitative and Quantitative Staffing Methodology” [EATCHIP, 1998c] and the “Report on Issues in ATCO Manpower Planning [EATCHIP, 1996]).

5.3.2 External / Internal Recruitment

A policy needs to be adopted which states the sources from which to recruit. An initial decision has to be made as to whether and to what extent

- future controllers should be recruited from internal staff resources (e.g. ATC Assistant [ATCA] or Flight Data Assistant [FDA] grades) and whether they should get an *ab initio* type or an adapted training;
- fully trained controllers should be recruited from external sources (e.g. from other States) and should receive conversion training or whether
- an *ab initio* approach should be followed or
- a combination of the foregoing.

In doing this consideration should be given to the needs and requirements within the operational environment, the training content and time, existing career strategies, State legislation, union agreement, etc.

5.3.3 Competencies and Training of Recruitment and Selection Personnel

Recruitment and selection involve tasks which can only be effectively done by staff who are either

- professionals or experts who provide expertise and experience in selection development, maintenance, control and validation;
- managers who have specialised in and are primarily responsible for HRM in the area of recruitment and selection;
- staff who have been trained to the level required to assist in the administration and/or application of methods in recruitment, in the application of procedures and tools in pre-selection or selection (e.g. paper sifting, administration of tests, interview board).

A clear policy should be available which precisely describes which tasks in recruitment and selection will be contracted out to service providers and which tasks are done within the ATSP organisation and by which level of staff.

A recruitment and selection manual should then enable all personnel involved in all stages in recruitment and selection to know their responsibilities, roles, tasks and competencies required. The manual will also describe what kind of training is required and the appropriate level at which certain competencies are required. Particular training will be necessary in methods and procedures to be applied in certain stages of the selection process (e.g. tests, use of assessment scales).

Chapter 5 of the selection guidelines document (EATMP, 2001b) details the requirements and gives recommendations with regard to quality assurance and control (training of selection personnel, working relations with contractors) which should be followed.

5.3.4 Re-testing Policy

A policy should be adopted if reapplication or re-testing of candidates is intended. It should precisely describe if and under what conditions applicants are allowed to reapply and should give exact rules and procedures (see also Sub-chapter 10.8).

5.3.5 Relocation Policy

OJT at some centres with high-density traffic is difficult for some students. Cases of failures can occur due to lack of progress in the training. The cost of those failures can be reduced by relocating those students to units with lower density of traffic. The above-mentioned trainees must obtain the check-out in the new unit. If this is not the case, the failure is confirmed.

Relocation may be applicable to a limited number of States and it requires setting up a system of clear procedures and decision-making. The low-density units do not always accept this solution of relocation. It should remain an

exception and should be applied only to very specific cases of failures, which occur at the end of the training. Most of all, it should not be a pretext to lower the norms of selection. One should avoid the relocation of several difficult cases in the same unit. The management of the relocation process should therefore be under the direct responsibility of the HR Managers.

A policy should describe if and under what conditions relocation of trainees should be allowed.

Guideline 18: Recruitment and Selection Policies

In accordance with the needs and requirements of the ATSP organisation, policies should be adopted which give clear guidance to staff by providing rules and conditions to be followed.

Chapter 2 of EATMP (2001b) details selection policies with regard to the selection plan, marketing and applicant information which should be followed For further information on marketing methods of the job, please see the report "Marketing and Communication of the ATCO Job - Phase 1: Study" (EATMP, 2000d).

5.4 Practical Implementation Strategies

Rather than guidelines the following sub-chapters are intended to provide a series of questions that should be considered when a selection system is set up.

The questions could be used as a checklist to ensure that everything has been considered. It should be recognised that not all questions will be applicable, as the situations will be different from recruitment to recruitment and State to State.

The starting point for the questions assumes that there is no structure in place to conduct the recruitment and selection and that there has been no previous recruitment and selection of *ab initio* trainees. The questions can be broadly divided into two groups:

- implementation strategies (i.e. which need to be in place before any of the process can take place) and
- selection management, the running of the process.

5.4.1 Manpower Planning

The starting point for any recruitment and selection process is a requirement for staff. This should come from an effective manpower plan. To draw up a plan the following should be considered:

1. Who should draw up the plan? Should it be done centrally or by each individual operational unit?

   If done by individual units: Should it be collated to ensure that vacancies and surpluses are balanced off?

2. What are the current and future manning needs, taking account of operational positions, relief, shift patterns, restriction of working hours, average sickness, training and leave (see also EATCHIP [1998c], the "Guidelines for ATCO Manpower Planning Processes [EATMP, 2000e] and the Report "Long-term ATCO Manpower Planning Simulation (LAMPS) – User Guide" [EATMP, 2001e])?

3. What are the known retirements over the period of the plan (e.g. the next five years?) - What are the losses due to medical reasons, loss of licence due to reduced efficiency, and general resignations?

4. What is the training time from *ab initio* to operational checkout? A plan that shows a requirement for controllers in two years while the training time is three years is clearly not viable.

5. The manpower plan represents the required number of controllers to be delivered to the operational system at particular times. These are the output numbers from the training system. For this plan to be of any use to recruitment, additional factors need to be taken into account:

   - What are the losses during training, college and OJT due to failure, resignations, etc.?

   - What are the possibilities of changing to a career other than controller?

6. What are the capacity constraints in Initial Training and OJT?

The addition of these figures produces an input plan and, if a staff requirement exists, then there is a recruitment need.

5.4.2 Recruitment

Once a recruitment need has been established then the following broad issues need to be addressed:

1. Should the recruitment be done internally or be outsourced to a recruitment or other ATC agency? In whole or in part? (e.g. advertising, response handling and testing could be outsourced but interviewing and contract administration kept internally).

   If outsourced: Who will monitor the process to ensure that it will meet and deliver the required numbers on target?

If to be done internally: Who should do it, a specific recruitment section, the HR department, the training department or operations? Who will be accountable and who will be responsible? How many staff and at what level will be required, full time or as part of other duties?

Staff numbers will of course be very much dependent on the recruitment target, the expected number of applications, etc. Should there be regional offices to handle initial parts of the selection or the whole selection process? Again numbers and the geography of the State concerned will be a key issue here.

Note: Management of the process will be difficult if not impossible without clear definitions to the above.

2. Will a formal quality management system be adopted for the recruitment and selection process, e.g. International Standards Organisation (ISO) 9000?

3. What experience and training do staff need, i.e. trained psychologist, HR expert, general administration, controller?

4. How will application data be processed, manually, by computer, etc.?

5. What equipment is needed: computers, optical readers, data store, etc.? What software is needed? What kind of database is needed to monitor the whole process?

6. What accommodation is needed?

   Note: The size of office accommodation required will be dependent on the number of staff and the systems they will be using to conduct the recruitment, e.g. computer workstations, filing cabinets, etc.

7. Should the recruitment be for *ab initio* controllers or can the need be met by recruiting experienced controllers from other agencies either from within or external to the State?

8. How will the applicants be attracted? (e.g. using adverts, newspaper, magazines, cinema, television, radio, the Internet, poster campaign, careers exhibitions, schools visits, etc.?

   Note: The need to advertise is linked to the economic and employment situation in the State concerned.

9. Should there be a recruitment brochure detailing the career of an ATCO and the recruitment and selection process?

10. What terms and conditions will be applicable to the students: employees paid whilst training, student grants, paid training plus pocket money, loan recovered from salary, self-funding?

Note: This again will be very much dependent on the climate and economic situation.

11. What policy will be adopted for incomplete application forms, instant rejecting, contact applicant for more information?

12. Will reapplications be allowed and, if yes, how many times, immediately or after a fixed period of time?

13. How long should rejected applications be stored before destruction?

5.4.3 Selection

Before selection can take place the following should be considered:

1. Will an application form be used or will application be by Curriculum Vitae (CV)? What information will be collected by the application form if used?

   Note: Legal requirements may dictate that some information may not be collected.

2. How will a paper sift be conducted, by whom, against what criteria, and what score mechanism will be used?

3. Are ability tests and personality questionnaires to be used? Are they to be “off the shelf”, specifically designed for controller selection, either in house or through an external agency? Should external experts or consultants be appointed to manage the selection process, while encouraging them with performance-related bonuses (e.g. present selection outputs can be ascertained and target outputs can be rewarded)?

4. Will tests be paper-and-pencil- or computer-based or both?

5. If ability tests are to be used: Should there be a dedicated test and interview centre or will the system be portable and allow regional administration?

   Note: The cost equation may depend on the answer to 7. (see below). If a large number of candidates are being processed it may be more economical in the initial stages to have regional test centres.

6. Will candidates be given practice materials prior to attendance at test centre to ensure that all applicants have an equal chance?

7. Will candidates be paid travel and subsistence expenses to attend for test and interview?

8. Will there be an interview, and if so, what style, biographical, situation, structure? Who will conduct it, panel or one-to-one, controllers, HR specialist or psychologist or both?

9. Will a cut-off score be applied for the tests or will selection for interview be top downwards?

   If cut-off score is used: Is it fixed or variable depending on the number of applicants?

10. What criteria will be used at interview and what scoring system will be used? Will selection for training be top down or will there be fixed criteria?

11. Who will make the final selection decision: interview panel, HR, independent assessor, etc.?

12. Will feedback be given to applicants on their results following the selection process?

    If feedback is given: Is it done automatically or on demand? In what timescale? Is feedback given to all applicants or only after certain stages?

13. Who will conduct the security clearance, if required and take up references?

14. At what point will the medical examination be conducted, at candidates' expense prior to application, after interview, etc.? Is prior medical information required from the candidate, e.g. eyesight prescription / certification?

5.5 Selection Management

The following questions relate to the actual process of recruitment and selection and are concerned with the mechanics rather than the policies.

5.5.1 Manpower Planning

1. Has a channel for communication been set up between MP and the selection department, regular meetings, updates in the manpower forecast?

2. Are manpower planners aware of constraints on recruitment and selection and taking them into account in their forecasts, e.g. recruitment timescales, the economic situation, etc.?

3. Has the manpower plan been authorised at a senior level or is it just an operational unit wish list?

5.5.2 Recruitment

1. Is there a need to advertise or is there a sufficient number of candidates applying from other marketing strategies?

2. Is there a system in place to monitor and feed back the effectiveness of the marketing strategies, e.g. specific reference numbers for each advert, question on application form "How did you learn about the vacancy", etc.? See also EATMP (2000d).

3. Is the response sufficient or is there a need for other marketing strategies?

4. If application forms are used: Will initial request for a form be by telephone, letter, tear-off slip, the Internet?

5. If telephone response is used: Will an initial telephone sift be conducted against fixed criteria, e.g. age, education, etc., or will all enquirers be sent an application?

6. Is equal opportunities monitoring to be conducted, e.g. gender, ethnic grouping and marital status?

   If so: Is a monitoring form available, is it separate from the application form, how will the information be recorded and by whom?

7. Are application forms being recorded on receipt, allocated a reference number, is data being extracted and recorded effectively?

8. Is there an appropriate filing system for the application forms, by reference, or name or region or combination of the above?

9. Are applications to be acknowledged?

   If so: In what timescale (e.g. days of receipt, three days, five days)?

5.5.3 Selection

1. Are the paper sift criteria clearly stated and available to all those who will conduct the sift? Is there a system in place to record the paper sift decisions? Is there a check to ensure consistency of decision-making?

2. Are applicants informed of the paper sift decision within a reasonable timescale? Are rejected applications filed and destroyed in accordance with policy? Is a system in place to monitor this?

3. If paper-and-pencil tests are used: are there sufficient test materials available, booklets, answer sheets, etc.?

4. Have sufficient test and interview centres been booked?

5. Are applicants given sufficient information and time to prepare for the tests and/or interview?

6. Is there a system in place to record test and interview results? Are candidates informed of the results within agreed timescales?

7. Are there procedures to deal with applicant complaints, e.g. non-selection, non-compliance with system?

8. If there is a quality monitoring system: are all checks being carried out and recorded?

9. Are interviewers fully trained and current?

10. Are test and interview procedures documented to ensure standardisation? Are the interviews being conducted consistently?

11. Are there procedures to conduct medical, security clearance and references?

12. Are candidates given sufficient notice of training start dates and enough information on the training course and its requirements?

5.6 Cost and Benefit Issues in Recruitment and Selection

The assessment of costs and benefits of applied procedures and tools in the selection process aims at identifying the impacts of proposed options in selecting candidates.

Note: This Chapter outlines the general procedures that should be followed and gives examples on the issues to be taken into account in identifying costs and benefits. The Chapter does not aim at outlining a methodology of an appropriate Cost-Benefit Analysis (CBA). See also EATCHIP, (1998b).

The selection of *ab initio* trainees is based on

- strategic planning of quantitative manpower requirements (e.g. succession planning);
- planning and implementation of recruitment campaigns (e.g. right timing, media to be used);
- the number of controllers needed within a certain time period;
- legal and/or organisational obligations in recruitment and selection;
- qualitative requirements in the operational environment of controllers (e.g. tasks to be fulfilled, training needs, required level of competency);

- training procedures and schemes to be applied;
- training capacity (e.g. capacity in Initial Training, OJT training capacity);
- planned future evolutions of the ATM system (e.g. automation, traffic increase);
- business strategies and objectives linked to HRM (e.g. career plans for operational staff, required workforce potential, required workforce flexibility).

These are only some considerations drawn from the viewpoint of an integrated HRM system which aims at putting people-related issues into the context of broader business requirements.

Because of the fact that differences in all or some of these related issues exist between ATSPs, CBAs need to be customised as is explained in the following example.

Example: The early identification of certain capabilities in selection will possibly include impact assessments concerning loss of otherwise suitable candidates, costs involved in development, application and maintenance of specific selection instruments, resources needed and the actual costs and benefits from applying these instruments.

The following suggestions and recommendations concentrate on those issues that are directly or indirectly related to recruitment and selection of *ab initio* trainees.

Guideline 19: Integrated Approach in Recruitment and Selection

Recruitment and selection should be in line with strategic plans and objectives and should take into account environmental, technological and structural changes together with desired future culture of the organisation and the likely impact on job and people profile.

5.6.1 Recommendations and Suggestions

Estimating the Benefits / Potential Cost Reductions

Cost reductions do not necessarily contribute to the overall benefit sought, i.e. higher success rates after training. An example of this is lower test administration costs that might result in lower reliability of the testing.

However, there are areas to be considered as potential cost reductions that may contribute secondary benefits or maybe even elements that are initially to be considered as costs but could end up as overall benefits.

Notwithstanding the above, the following are examples of the potential benefits / reduction in costs that could accrue from enhanced selection methods:

- reduced number of ability tests of high reliability and validity (e.g. established on the basis of a validation study);
- higher number of interviewees with higher potential in the interview (e.g. by enhanced ability tests, or improved cut-off scores);
- higher reliability and validity of the interview (e.g. through enhanced training of interviewers, improved interview procedures);
- higher quality of the candidates that enter training (e.g. better match of person / training and person / job profiles).

Estimating the Costs

As in estimating the benefits, first it is necessary to list the items which might directly result in expenditure with the implementation of a test battery. The following list is not exhaustive, however such costs might include:

- personnel costs for staff in selection departments;
- administrative overhead costs;
- costs of research and development support (e.g. test maintenance, test development);
- hard- and software lease or purchase;
- hard- and software installation;
- training of selection personnel and interview board members;
- costs of on-site test administration (e.g. personnel costs, accommodation);
- annual updates (e.g. manuals, norms);
- costs of regular reports.

Finding a Balance

Which of the assessed options is selected depends, primarily, upon the cost and efficiency benefits to the user and, secondly, on the alternative to allow further steps to be taken towards the establishment of an enhanced selection system.

Having generated estimates, based on the best available information on costs and benefits, it is necessary to merge them to find a balance and to arrive at decisions.

Usually benefits will be translated into direct financial gains by 'a per year saving' compared to current costs and are viewed over a period of several years (e.g. in a five-year plan).

However, financial (quantitative) benefits or savings in terms of money and resources are not the only performance indicators of an efficient and effective recruitment and selection system. Benefits as well as risks in regard to important qualitative aspects will be based on judgements rather than exact figures.

Guideline 20: Qualitative Cost and Benefit Issues

Qualitative aspects in terms of the desired organisational culture and the business perspectives and requirements, the requirements concerning workforce flexibility and mobility, future career perspectives of air traffic controllers, etc., should as well be taken into account in arriving at a decision concerning selection methods and tools taking into account the costs involved.

Page intentionally left blank

6. RECRUITMENT ISSUES

This Chapter outlines influencing factors and some of the ways in which external recruitment of *ab initio* trainee controllers can be done in ATSP organisations.

Details on recruitment and pre-selection strategies are given in (EATMP, 2001d).

6.1 Guiding Principles in Recruitment

Selection is contingent upon successful recruitment (Hopkin, 1995). Recruitment aims at attracting individuals and making them interested in applying for a job in ATC.

The more interested and qualified the applicants, the more sophisticated and valid selection procedures applied in later stages can in principle become (Hopkin, 1995).

Recruitment involves therefore the timely search and identification of high calibre applicants in sufficient numbers so that the most appropriate can later be selected. Hence, recruitment aims to increase the success rate in subsequent selection and training by attracting applicants, which have both the right potential and motivation for controller training and job. Recruitment aims as well at increasing the chance of retaining candidates once they are hired (Schuler and Huber, 1993).

High numbers of applicants call for a multi-stage selection process with, for example, an initial pre-selection of applicants based on mandatory basic requirements, possibly a paper sift stage based on identified characteristics in an application form and further stages comprising various selection procedures.

6.1.1 Recruitment Activities

The following are major activities in a systematic and effective recruitment programme and will be tackled in more detail in Sub-chapter 6.2, “Policy Recommendations”:

- establishing recruitment needs in conjunction with the unit(s) responsible for MP and training;
- increasing the pool of quality applicants at reasonable costs by applying effective marketing procedures and choosing appropriate and effective advertising and job promotion media;
- staying informed about and reacting to the conditions and possibilities of the job market;

- developing effective and informative recruitment material;
- recording the numbers and the qualities of applicants hired or not hired from various recruitment sources for later evaluation and validation;
- accomplishing all recruitment activities within the framework of legal and social obligations.

6.2 Policy Recommendations

6.2.1 Planning of Recruitment Needs

Recruitment is the first step in the wider HRM process in ATC and is closely linked to MP.

The aim of MP is to provide a sufficient number of qualified controllers, on a timely basis, to ensure the provision of ATS within a given environment. This can only be achieved if close consultation between MP, training authorities, recruitment and selection is ensured [2].

Manpower planners must take into account the

- time needed for advertising, sending in applications, registration and acceptance of applications, pre-selection of applicants and final selection;
- success rate in different stages of recruitment and selection;
- time needed for medical and security checks;
- time to be allowed for selected candidates to enter training;
- training capacity in different stages of training and in particular in OJT;
- total length of training;
- success rate in training;
- average time for achieving the necessary qualifications (ratings) at operational positions.

A continuous, uninterrupted recruitment, selection and training of *ab initio* trainee controllers should be the aim.

[2] For further details see EATCHIP (1996), the “Human-Machine Interface Catalogue for EATCHIP Phase III – General Principles” (EATCHIP, 1998d) and EATMP (2000e).

Guideline 21: Planning of Recruitment Needs

Recruitment needs must be established at an early enough stage to ensure that recruitment, selection and training activities and capacities can be appropriately planned taking into account the time, capacity and success in these stages.

6.2.2 Advertising Media and Promotion of the Job

Although the evidence on the effectiveness of advertising media shows that there is no definite answer or 'best' solution to it the following general policy applies:

The direct identification of individual characteristics of applicants that are related to success in training and in the job is of ample importance. It enables the organisation to selectively utilise recruitment sources and media according to their ability to

- tap the desired characteristics of applicants / the target population,
- distribute relevant information on these characteristics,
- develop a selection process which increases training and job success.

Not all advertising media are appropriate and efficient in this respect. This should be established by a market research on local and national advertising media.

It must be kept in mind that different media are accessed and used by different groups of people and that it is not only the number of applications gained from a source but the quality of applications which is important.

An address where those attracted by the job can obtain detailed information should be given. In addition to this toll-free telephone numbers and/or World Wide Web (WWW) site address can be given in the advertisement to provide applicants with further information (see also Annex C in EATMP, 2001d).

Guideline 22: Advertising Media and Promotion of the Job

The choice of a source for recruitment should depend on the capability of the media to tap and give information on desired characteristics of applicants and to address the population, which to a great extent possesses these characteristics.

6.2.3 Observation of Job and Labour Market

The job and labour market entered for external recruitment should be regularly and closely followed and observed in order to

- find out what is the appropriate timing of recruitment procedures (e.g. end of school year, end of military service) to enhance the chance of obtaining "good" candidates through early entry into the recruitment process;
- take into account and react to any significant changes in the job market (e.g. changes in the unemployment rate, total number of potential applicants due to age and/or education);
- identify significant regional differences in recruitment behaviour in terms of mobility, educational levels, good knowledge of the English language, economic focus, etc.;
- identify target groups of well-suited applicants and media to encourage individuals within target groups to apply.

This information can be used to better target appropriate applicant populations and to make decisions on appropriate dates, media and campaign strategy.

Appropriate timing of recruitment is also necessary to avoid delays for applicants in the process. Delays are viewed negatively by applicants and lead to negative impressions and decisions (e.g. taking other offers). Late market entries and delayed follow-ups lower job acceptance.

The goal in recruitment should be to attract applicants' attention early and keep it through prompt follow-ups and feedback.

It should also be kept in mind that a stop / start approach in recruitment (e.g. stopping recruitment for one year), a practice not unusual in ATSP organisations, quickly lowers applicant awareness and attention in the job market, which cannot easily be recovered.

Guideline 23: Job and Labour Market and Recruitment Planning

Job and labour markets should be closely observed and monitored. Important changes need to be identified and local differences detected to plan recruitment campaigns on a timely basis, to attract the most suitable population and to avoid unnecessary delays in the recruitment process.

6.2.4 Recruitment and Information Material

Appropriate information material about the ATSP organisation, the current training system and the controller job is one of the means of attracting well-suited candidates. This information should be sent to all applicants on their request once they respond to an advertisement.

There is evidence that ATSP organisations would be well advised to be as accurate as possible in describing the training including the success rates, the job and the work environment of controllers.

Although this strategy may cause some applicants to turn away from the job, available evidence shows that those who received realistic information tend to have longer organisational tenure, that is stay within training and later in the job, than those who received only favourable information.

Sufficient and realistic information is therefore a strong tool in the self-selection process of applicants and could include:

- descriptions of mandatory basic requirements,
- abilities and skills identified by selection tests,
- skills to be achieved in institutional and OJT,
- controller working conditions,
- important social attitudes of successful job incumbents,
- success rates in training.

Note: Detailed information on the administration of selection tests, as outlined in EATMP (2001b) should be sent to applicants which have passed the pre-selection stage and are invited for testing.

Guideline 24: Information and Recruitment Material

Recruitment material should contain sufficient, accurate and realistic information to attract well-suited applicants. It should be designed to help them make decisions regarding their suitability (against stated requirements) in selection, training, in the work environment and compared to characteristics of current job incumbents.

6.2.5 Evaluation and Validation of Recruitment Activities

In order to develop a systematic and effective recruitment programme it is necessary to follow up on the achievements or success of all applicants / candidates in:

- the later stages of the selection process,
- the later stages of Initial and Unit Training,
- the job.

This can be done by recording the numbers and the qualities of applicants from various recruitment sources and locations, different regions, different information sessions, etc., for later evaluation and validation purposes.

It should be apparent that data from recruitment could be evaluated and validated in the same way as test scores. Expected recruitment targets among sub-groups for example can be compared to actual results achieved.

The evaluation should, as far as possible, also include applicants not hired.

The data gathered should be included in the recruitment, selection and training database described in detail in Chapter 12.

Guideline 25: Evaluation and Validation of Recruitment Activities

Activities in Recruitment and the efficiency and effectiveness of methods and media used should be evaluated and validated on the basis of data against criteria in later selection and training stages.

6.2.6 Legal and Social Obligations in Recruitment

Recruitment and the supply of labour are to some extent regulated by national employment legislation and by the role of national employment offices. The breadth of potential legislation affecting recruitment is large ranging from codified legislation to national collective agreements or codes of best practice.

In the recruitment stage the following legal elements might need to be taken into account:

- sources of labour supply (e.g. recruitment from particular sources of labour as for example unemployed people, military ATCOs);
- discrimination and equal opportunities (e.g. with regard to minorities, gender, state of health, handicap) or affirmative action;
- recruitment codes of practice (e.g. content of application form);
- disclosure of information and security of personal data.

Depending on existing obligations in relation to the first two bullets, recruiters need to keep track of applicants and to record data to show proof on any affirmative action to be taken (e.g. by taking into account the number of past required applications for every one successful candidate within every designated group, for example females).

6.2.7 Professional Orientation and Career Counselling: An Outlook

Support to career counsellors will help to increase the flow of applicants for employment as ATCOs and promote the selection of candidates with a high ongoing motivation to enter the profession.

Career counselling could be part of a broader programme of career orientation helping young students to find their way in existing professions and make the right choice of their future career.

There is evidence showing the importance of early occupational orientation for professional motivation, stability and expertise in professions (job performance).

Page intentionally left blank

7. SELECTION TESTS

7.1 Introduction

In selection tests are used (in conjunction with other information, for example, from application forms or from interview) to make predictions about later performance of individuals in training and/or in the job, and to arrive at decisions whether an applicant should be accepted or rejected.

It is clear from Chapter 4, "Validation Issues", that these predictions and decisions should be validated.

This Chapter addresses the underlying notions of selection tests and the considerations that should guide implementation and practical application of these tools.

Tests refer to a great variety of published diagnostic, prognostic and evaluative devices of either the constructed performance or structured behaviour sample. The following sub-chapters provide basic information, which aims to have a fair understanding of the major categories of tests.

7.1.1 Differences Between Test Instruments

Performance Testing

Test instruments can be divided into three broad categories:

1. **Constructed Performance Tasks**: These are tests (e.g. conventional standardised ability tests), which are designed to isolate and assess particular psychological constructs (e.g. spatial comprehension) without actually simulating the environment (e.g. the ATC environment). The media in use for these tests are paper-and-pencil, computer and oral exchange.

2. **Questionnaires and Inventories**: These are tests (e.g. personality or interest inventories) through which people can be asked to report personal interests, attitudes and typical reactions to situations which might be faced on the job. The media used are paper-and-pencil or computer.

3. **Structured, Observable Behaviour Samples**: These tests provide standardised assessments of complex skills and working styles relevant, for example, to simulated ATC decision-making, to ATC communication skills, or to styles of performance under workload stress or in the face of complex airspace (after CDSEPT, 1999). These tests are mainly computer-based and require trained observers / assessors who make assessments during testing.

Note: In assessing structured, observable behaviour samples, as mentioned under 3 above, trained assessors (e.g. ATCOs, HR specialists, psychologists) will **assess** the performance or the behaviour of test takers as it is observed but they are not supposed to **evaluate** characteristics (e.g. clinical features) of applicants.

Tests usually include:

- standardised instruments to measure cognitive, perceptual or physical abilities, non-cognitive abilities and achievement in more complex tasks;
- interest inventories to measure vocational or avocational interests;
- personality questionnaires (or inventories) to measure personality traits or behavioural tendencies (CDSEPT, 1999).

The categories of diagnostic and evaluative devices (e.g. for clinical assessments) and projective instruments will not be further considered in this document.

However, ongoing research into cognitive psychology is producing new instruments using computerised testing, interactive testing and multimedia test presentations. In addition to this, physiological and neuropsychological measures are being investigated increasingly as potential selection and classification devices (CDSEPT, 1999); see for one example Sub-chapter 7.6.

7.1.2 Testing

Performance Testing

Testing refers to a measurement or assessment process whereby one obtains a quantifiable estimate of some aspect of current performance. Most tests consist of constructed tasks, i.e. specifically designed test items upon which performance is measured. An evident example is the task of answering a question.

An important distinction between tests relates to the measurement situation where there are two broad categories:

1. **Tests of Maximum Performance**: The test takers are expected or instructed to do their best in the test (e.g. in ability tests, intelligence tests, work sample tests, knowledge tests, achievement tests).

   This category of tests is further divided into:

   - **Speed Tests** where test takers are requested to work as fast (and accurate) as possible sometimes for a very short period of time. Some ability tests are **speeded** because a 'true' speed test consists of items which, if given unspeeded, would be correctly answered by virtually everybody. Examples of such items are computational speed, e.g. addition and subtraction of simple sets of digits (Kline,

1995). Typical test examples are concentration tests, perceptional speed tests and some tests in an intelligence test battery, etc. The performance measured is typically the number of correctly solved test items.

Most ability tests used in *ab initio* trainee selection are speeded tests.

- **Power Tests** where test takers are requested to solve more complex problems and find the correct answer for them. Power tests are usually untimed and unspeeded, where test takers are given as long as they need to complete the test. However, the overall time limits normally allow that all items of the test can be tackled. The performance in the test will depend on the knowledge and skills of test takers. Normally, the number of correct solved items is counted, taking into account guessing (in case of multiple-choice items; see Sub-chapter 10.6 for more details).

  Most achievement tests are power tests. This category of tests is also used in selection of *ab initio* trainees (e.g. English knowledge tests, tests for technical understanding).

2. **Habitual Performance Test**: The test takers are requested to report or otherwise indicate what they believe or feel, or how they would habitually act in a given situation (e.g. described in a personality questionnaire or inventory) (Tenopyr, 1987).

Norm-referenced Testing and Criterion-referenced Testing

Tests can be applied in different contexts:

- In **Norm-referenced Testing** the candidate's test performance is related to the standing in some reference group. For example, test takers who can in 60 seconds correctly identify which aircraft are in conflict among 8 moving aircraft targets specified by altitude, speed and route are above average in their performance compared to a reference group of the same age and education.

- In **Criterion-referenced Testing** the candidate's test performance is related to specific performance standards. These tests are usually interpreted in terms of mastery or non-mastery - for example, 60 seconds as the criterion to be achieved in the test described above, regardless of the reference group of age or education.

This Chapter will focus on the major test categories in use for the selection of *ab initio* trainee controllers, which are:

- ability tests,

- achievement tests,

- interest inventories,
- personality questionnaires / inventories.

7.2 Ability Tests

Ability tests measure the performance of persons in a defined domain of their cognitive, psychomotor or physical functioning (e.g. spatial comprehension). These functions are attributes that cannot directly be observed because they are actually theoretical constructs. Nobody can directly 'see' or 'touch' or otherwise directly observe 'spatial comprehension' for example. However, the constructs as, for example, the cognitive function of 'spatial comprehension' have been derived from a theory which explains the construct and gives hypothesis regarding the differences between people in the outcomes in tasks where the cognitive function is used as a resource. The theory might state, that people who differ in their 'spatial comprehension' ability will show a difference in their performance in a test which aims to measure this ability.

Such a test would be developed in accordance with the theory on 'spatial comprehension' and might, for example, include items on left-right differentiation at different levels of item difficulty.

It needs to be borne in mind therefore, that the general aim in using ability tests in selection is, to **predict** future performance of the persons in other (more complex) **tasks** (e.g. radar simulation) where the ability is supposed to be used as a resource.

7.2.1 Difference Between "Ability" and "Achievement"

The definitions and the vocabulary used in testing can cause problems in understanding what is actually meant. This is in particular evident with regard to "ability" and "achievement":

- **Ability** as mentioned above refers to a psychological **construct**. The difference between ability and achievement can be understood best by an example: "Spelling" is not an ability, but a skill that has been achieved. However, the acquisition of the skill how to spell depends on certain abilities.

  The aim in selection of *ab initio* trainees is, to measure those abilities that are necessary for acquiring the skills that have to be achieved during training. If the test is used for **prediction** purposes it will be called an **ability** test.

  Note: As a psychometric concept the term **aptitude** is highly unsatisfactory and confusing. It embraces both abilities and attainments. As a scientific concept it should be abandoned (Kline, 1995). Therefore, in the following we shall only refer to tests as either ability or achievement tests.

- **Achievement** (or attainment) refers to the possession of **skills** which have been acquired, through experience and/or deliberate training. Achievement tests therefore assess how much has been learned whereas ability refers to the potential for acquiring the necessary skills.

  Applicants for *ab initio* training are tested, for example, with regard to their achievement in English or mathematics; skills they have acquired before. If the test is used to reflect past events (e.g. to test learning skills or the degree of mastery in a specific content area) it will be referred to as an **achievement** test.

  There are, however, exceptions: Although, for example, 'arithmetical ability test' requires some past learning, it is still mainly denoted as an ability test since it usually also predicts later learning success (Tenopyr, 1987).

  Note: It should be borne in mind that it is not necessarily the case that applicants with known past achievements (e.g. in previous ATC military training) will pass the ability tests which are applied to predict performance in ATC (e.g. civil ATC training).

7.2.2 Guiding Principles of Ability Testing

As mentioned in Chapter 4 on "Validation Issues", ability tests are among the best-known predictors of training and job success. They predict these measures better than other traditional selection methods, such as interviews and education evaluations. Ability tests predict training achievement better than job performance (Tenopyr, 1987).

"Cognitive ability has a mean validity for training success of about .55 across all known job families" (Hunter and Hunter, 1984).

7.2.3 Basic Assumptions

Two different rationales or assumptions are basic to ability tests and are still under discussion within the 'scientific community':

- **The 'Classical' Assumption**: The assumption is that individuals possess an innate capacity or potential which can be measured by ability tests. It is assumed that this capacity is to a large extent independent of the learning history of an individual.

  One of the derived assumptions in the classical approach is that individuals can only to a limited extent improve their performance in a certain area (e.g. in spatial comprehension).

- **The 'Modern' Assumption**: The assumption is that individuals have developed their abilities in certain areas (Anastasi, 1988). What ability (and in particular achievement) tests actually reflect are the effects of experience or, simply, the effects of learning (Reschly, 1990).

General intelligence underlines many human abilities. Annex A of EATMP (2001d) refers to the 'g' or general intelligence. Although the 'g' (primary) factors account for a large proportion of the measured variance of ability, other factors (which are not completely measured by intelligence tests) may have high face validity to the ATC job. The 'g' factors are necessary to predict success in most jobs, but the specific abilities may also be required.

Among the mental abilities the following primary factors are important:

- verbal ability (e.g. understanding words and ideas, verbal analogies and verbal reasoning);
- spatial ability (e.g. visualisation and recognition of Three-Dimensional (3-D) geometrical shapes when they are reoriented in space);
- numerical ability (e.g. manipulating numbers speedily and accurately);
- word fluency (e.g. producing as many words as possible with a given beginning, as in the game 'Scrabble').

Some abilities are considered as 'second order' abilities since they result from the correlations among the primary factors group.

For example, the second order factor called 'cognitive speed factor' is a minor factor in a numerical computation problem ('numerical factor' is a primary factor).

A retrieval capacity or general fluency can be measured by 'ideational fluency', association, or by producing many ideas, or original uses on any subject. However, 'ideational fluency' is a primary ability factor known as 'FI'.

'Second order' abilities can sometimes be regarded as highly specific. Cattell (1973) called them 'bloated specifics'. However, these 'second order' abilities can still be most useful in personnel selection.

Typical and relevant ability areas where candidates for *ab initio* trainee controllers are tested therefore include specific abilities:

- memory functions,
- attention,
- logical reasoning,
- mental arithmetic,
- spatial comprehension,
- perceptual speed.

For further details see Sub-chapter 4.3.1 of EATMP (2001b).

In addition to this

- perceptual / psychomotor abilities such as motor coordination, multi-limb coordination, arm-hand steadiness, reaction time tests, or electrical contacts register tests

could be considered in *ab initio* selection.

Guideline 26: Ability Tests and the Test Battery

Ability tests covering main cognitive (mental) and psychomotor abilities should be included in the test battery (together with other elements such as a personality questionnaire).

7.2.4 Principles of Development of Ability Tests

Ability test refers to a collection of operational definitions of abilities which happen to be of value in a particular culture (Kline, 1995). Essentially an ability test consists of a group of factors and an arbitrary collection of skills. For example, the widely used Differential Aptitude Test (DAT) (also for selection of *ab initio* trainee controllers) includes unifactorial tests of primary abilities (e.g. verbal ability) and multifactorial, or collection of skills (e.g. clerical abilities which involve speed, accuracy, conscientiousness, etc.).

The difficulties with multifactorial ability tests, such as the DAT, is that two identical scores do not necessarily have the same psychological meaning. In addition, ability tests tend to be saturated with the 'g' factor (see also Sub-chapter 3.1.2 of the TS).

7.2.5 Practical Application

Relevance for Controller Job: Ability tests are often negatively correlated with training time, i.e. good performance of a candidate in an ability test is usually related to subsequent short training time.

A statement in a technical manual which may suggest that spatial ability is required for *ab initio* trainee controllers may quantitatively be inadequate, because it does not indicate the degree to which a controller's success has been found to depend upon spatial ability.

Social Desirability Responses: Ability tests cannot be affected by misrepresentations of the socially desirable manner.

7.2.6 Policy Recommendations for Ability Testing

Recommendations for Implementation: When ATC training is designed to start at a low level of mastery, it would be appropriate to use ability tests which are supported by a criterion-related validation study for the relatively untrained candidates. In the selection of *ab initio* trainee controllers to training in a field in which they have no experience, the measurement of a differential ability is

indispensable. It will be much more useful as a sensitive index than achievement tests.

If, for some reason, ATC training is to start at a higher level of mastery and applicants are selected on the basis of their current level of learning, a higher weight for achievement tests may be more appropriate (Tenopyr, 1987).

Candidates for *ab initio* training are not supposed to master specific ATC procedures, but rather have the capacity to be trained for this.

It makes sense to administer tests measuring general ability in addition to ability tests, since it can be assumed that if two persons have acquired in training the same level of skill, the one with the highest 'g' score is also more likely to apply the knowledge or to acquire new knowledge in the future as ATC progresses (after training).

Ability or Achievement Test? It should not be assumed that, just because the word 'ability' (or sometimes 'aptitude') is used in the title of a test, it measures a construct distinct from what is measured by an 'achievement' test.

Selection decisions based on the differences between ability and achievement for a candidate for *ab initio* trainee controller should take into account the overlap between the constructs.

Guideline 27: Ability Tests

For most situations in which candidates are selected for training, it is advisable to use ability tests rather than some other selection methods.

Unifactorial tests should be used in preference to multifactorial scales wherever possible. Separate abilities are best measured separately.

7.3 Achievement Tests

The guiding principles and basic assumptions of achievement tests have already been mentioned before (see Sub-chapters 7.2.1 above).

7.3.1 Principles of Development of Achievement Tests

Achievement tests are often constructed to be applied in a criterion-referenced testing context (see last part of Sub-chapter 7.1.2 above, "Norm-referenced Testing and Criterion-referenced Testing").

When achievement tests are developed to be criterion-referenced, (i.e. to indicate the candidates' performance relative to what **they** can do), as compared to norm-referenced (i.e. to indicate the candidates' performance relative to what **others** can do), classical reliability theory may not be applicable and special reliability formulas may be required (Tenopyr, 1987).

Subkoviak (1984) describes some of these 'threshold loss function' formulas and other special reliability considerations in achievement tests.

7.3.2 Practical Application

Relevance for Controller Job: The fact that fully licensed controllers possess certain knowledge or skills is very relevant, but not persuasive.

For example, basic knowledge of aeronautical meteorology or origin and characteristics of weather phenomena affecting flight safety may be required in a licensing exam according to Annex 1, Chapter 4.4.1.2 (§f) of the International Civil Aviation Organization - Personnel Licensing (ICAO, 1988).

The claim that a particular skill is necessary for ATCOs should therefore be supported by evidence, such as job analysis. It provides the reasons why specific skills or knowledge are required from ATCOs.

An obvious area of skill which should be tested in *ab initio* selection process is English language. Another area that might be considered is mathematical knowledge.

7.3.3 Policy Recommendations

English Language Tests: A classical attainment English test will include grammar, vocabulary and syntax. English language proficiency should not be determined solely with tests that demand only a single linguistic skill (e.g. vocabulary). A more complete range of language skills (e.g. spoken level) needs to be assessed (CDSEPT, 1999).

With regard to the ability to speak English, observing a candidate's speech in specific job situations and non-job-related situations could provide additional information about their proficiency in English. This may not, however, always be sufficient to judge solely their ability to express themselves in English without accent in formal radio communications. A fortiori, judgements on English fluency based solely on observing a candidate's speech in non-job-related situations (e.g. interview) may be inappropriate.

The (English) language proficiency level of the selection test for *ab initio* air traffic controllers should be equivalent to the language proficiency requirements specified by ICAO (Level 3 – Pre-operational) in the Appendix to Annex 1 Personnel Licensing, ICAO Language Proficiency Standard).

ICAO proposed amendments to Annex 10, Volume II, Chapter 5, paragraphs 5.2.1.2.1 and 5.2.1.2.2 specify the use of language(s) in aeronautical communications.

Mathematical Knowledge: Although tests in mathematics usually put an unwarranted premium on certain qualities - speed of recall, mental arithmetic - while ignoring hard-to-measure ones like creative thinking, they can still be

used as an evaluative device and can indicate a mediocre performance of candidates for *ab initio* trainee controllers.

Use in Validation: Achievement tests can be used not only for selection purposes but also for validating selection tests, diagnosing training needs or simply for evaluating trainees' achievement.

Legal Aspects: The user of achievement test results where different cut-off scores exist for different groups (e.g. age or gender) will be well advised to prepare some good evidence that differences in selection test scores are indeed reflected in differences in job or training performance. Members of affected groups in some States may be entitled to legally challenge the test. However, many achievement tests indicate different levels of achievement for different groups (Tenopyr, 1987).

Objectivity, Reliability and Validity: A good achievement test is reliable and valid, and covers the content of the courses or the knowledge, with many items within the full range of difficulty level, which are objectively scored (Kline, 1995).

Test Answer Categories: The most commonly used achievement tests are of multiple-choice items (e.g. a problem or a question followed by four-five options).

In a good multiple-choice test, items should be simple with only one correct option among equally plausible incorrect answers (distractors). Items should not be too trivial, obvious or sloppy.

Items which consist of statements where the test taker has to mark true or false increase the possibility of correct guessing to 50% (see also Sub-chapter 10.6).

Guideline 28: Multiple-choice Items

Multiple-choice items are more effective than true-false items in coping with the effects of guessing on test effectiveness and therefore it would be more advisable to prefer multiple-choice items to true-false items.

Guideline 29: Achievement Tests

Inferences made from achievement tests about future job performance are best supported by criterion-related evidence.

7.4 Interest Inventories

7.4.1 Guiding Principles of Interest Inventories

Interest is defined as an explanatory construct inferred from observations of a person's behaviour (Kline, 1995).

Inventories are instruments for identifying interests and personality characteristics through self-report (CDSEPT, 1999).

Inventory is a way to help people confront what they already know about themselves. This is the most convenient way of ascertaining what the interests of a person are. It is a more convenient way than the interview to present standardised indirect questions, which can be objectively scored (Cronbach, 1990).

Primarily interest inventories were developed to inform the respondent and not for selection purposes. Increasingly, interest inventories are also used as an instrument for predicting tenure in employment.

A well-known interest inventory was developed by Holland (1985), the Vocational Preference Inventory, which measures six variables of interest (i.e. realistic, intellectual, social, conventional, enterprising and artistic interests) and five other scales. It includes 160 occupational titles to which individuals have to indicate 'like' or 'dislike'. The theory claims that each variable defines a personality type and that people of a given type are attracted to jobs which suit them.

7.4.2 Objectives

Interest inventories are used to identify those candidates most likely to complete the training programme and stay in the profession (Reeves and Booth, 1979).

7.4.3 Basic Assumptions

Interest assessment is based on the assumption that occupational adjustment can be enhanced if a person's characteristics and interests match the requirements of the job. This is an additional piece of information, which is not provided by ability tests, for example.

7.4.4 Principles of Development of Interest Inventories

In the development of occupational interest inventories typically a wide range of activities is subsumed under a given occupational title. The extent to which average patterns of interests (e.g. reading aviation magazines) or abilities are compatible with the major specialities within the controller job should be reported to the users.

Directions of interest inventories should specify whether the test takers should mark the activities they would ideally like, or whether the inventory is intended only to cover realistic **activities** which candidates also had or have the opportunity and ability to exercise.

Interest inventories, which are designed to be answered by candidates, may stress the need for the help of professionally trained persons in interpreting their results.

Items in interest inventories should differentiate among groups (e.g. persons with similar or dissimilar interests). Items should also be gender fair; no item should suggest that ATC activity is more appropriate for one gender than the other. Items should also be culture-fair and easy to read (Hansen, 1990).

7.4.5 Practical Application

For interest and personality inventories (for the latter see Sub-chapter 7.5) intended for selection, evidence should be presented on the extent to which scores are susceptible to an attempt by candidates to present false or unduly favourable pictures of themselves (CDSEPT, 1999).

It is important to recognise in providing scores on all occupational scales in a specific interest inventory that not necessarily all scales have both male and female group norms.

7.4.6 Policy Recommendations for Interest Inventories

A manual of a controller interest inventory may mention the fact that even if a candidate has a high degree of interest in ATC, compared to the general population, this degree of interest can still be lower than that of a fully licensed controller actually engaged in ATC.

Practicality: Cronbach (1990) identified some of the disadvantages of interest inventories. In many cases, inventories refer to typical situations and not specific situations. Some of the responses require interpretation. Inventories also tend to be quite lengthy. Above all inventories tend to award high scores to conformers.

Cultural or Gender Bias: Interest inventories should be evaluated periodically because societal changes can make items obsolete (Hansen, 1990).

It is important to keep in mind that choices of interests and abilities may be influenced by cultural factors such as traditional sex-role social expectations and early experiences of various gender or ethnic groups. For example, female candidates for *ab initio* trainee controllers may have had fewer arithmetical-related experiences.

Age Specific: Caution should be exercised in interpreting interest inventories or ability tests when normative data were based on specific age. Although

these measures may still be useful with older candidates for a*b initio* trainee controllers, their scores should be interpreted cautiously.

7.5 Personality Questionnaires (or Inventories)

7.5.1 Guiding Principles of Personality Questionnaires

Personality questionnaires consist of a set of items (e.g. statements about feelings, wishes, attitudes, beliefs, behaviour, etc.) to which test takers have to respond (e.g. by agreeing or disagreeing with the statement). This tool can measure factors such as, for example, anxiety (Kline, 1995).

One of the techniques commonly used for personality assessment is a projective test. The projective test is a stimulus (usually ambiguous) to which candidates have to respond, designed so that it encourages them to project onto their own innermost thoughts, feelings, desires and emotions (Kline, 1995).

The use of projective instruments is often constrained by the nature of the open-ended responses, which are typical of these measures.

An example of a projective test is the Rorschach test, which is in use in Hungary for the selection of *ab initio* trainee controllers (EATMP, 2000a). This test consists of symmetrical ink-blots which candidates have to describe.

In other forms of projective tests, candidates can be asked to complete sentences or draw certain objects.

Another, more commonly applied form of personality assessment in addition to personality questionnaires is an objective test. Objective tests are defined as tests whose purport is hidden from subjects (Cattell and Warburton, 1967), for example, measuring the time taken to complete the task, when it was not the specified purpose of a test.

7.5.2 Objectives

Although the objective of personality assessments to increase the understanding of a candidate by providing a more complete picture of a person, from a professional view the value of personality inventories as predictors on the job has generally been pessimistic (Cronbach, 1990).

7.5.3 Basic Assumptions

The basic assumption to be made is that the behaviour of candidates in various stress situations, as may arise in an ATC environment, for example, may contribute to the general understanding of candidates as a person and thus to the prediction of their behaviour. It is assumed that personality questionnaires can contribute to this understanding.

7.5.4 Principles of Development of Personality Questionnaires

Among the techniques which are in use for the development of personality questionnaires, the factor analysis of the personality sphere has to be mentioned. This technique was behind the famous Sixteen Personality Factor Questionnaire (16PF). Cattel produced sixteen factors from rating of all trait terms - factors embracing the semantic personality sphere (Kline, 1995).

There are other non-factor analytic approaches for the development of personality questionnaires such as criterion-keyed tests. In this approach sets of items, which discriminate individuals of different personality, are developed. Some of the most commonly known personality questionnaires were developed by using this non-factor analytic approach, for example, the Minnesota Multiphasic Personality Inventory (MMPI) and the Californian Psychological Inventory (CPI). The CPI is an adaptation and extension of the MMPI. In the case of the MMPI the items were constructed from a knowledge of the typical symptomatology of patients seeking psychiatric help. Although originally constructed for clinical assessment, the MMPI is widely used for selection.

An EATMP survey identified 29 countries that use personality assessments for the selection of *ab initio* trainee controllers. It also revealed that 19 European ATSP organisations use personality questionnaires for that purpose (EATMP, 2000a).

7.5.5 Practical Application

Some of the disadvantages of personality questionnaires are as follows:

- Personality questionnaires often tend to be psychometrically powerful and reliable, but seldom valid. Establishing predictive validity and even face validity of items is difficult. In fact, demonstrating construct validity is almost the sole alternative.

- It has proved to be quite difficult to draw up an agreed list of variables which will account for a normal personality variance and which could be included in personality questionnaires.

- Item writing is difficult and the avoidance of simplistic, transparent statements when trying to describe the rich subtlety of human feelings is complex (Kline, 1995).

- When items are somewhat vague, test takers tend to agree with them regardless of their content.

- The endorsement of comprehensible, unambiguous items included in personality questionnaires is exposed in particular to the influence of keeping in line with socially desirable responses.

- Test takers are attracted to a middle, uncommitted option when one such is presented.

- Test takers prefer extreme responses on a scale, just because the responses are extreme.

- Cultural norms define roles and shape styles of behaviour and feeling.

- A response may depend on the immediate situation encountered on each particular occasion. Personality dimensions can be subdivided by categorising situations (Cronbach, 1990).

In order to overcome some of these difficulties, personality questionnaires use items such as true-false, yes-no, like-dislike and other item forms. To avoid response sets, the number of items keyed 'yes' and 'no' is approximately balanced.

Many personality questionnaires use 'lie scales' - introduction of trivial offences which almost everybody would normally commit - to detect social desirability patterns.

Even if efforts were made to construct 'good' items in personality questionnaires, the measured traits cannot easily been recognised. Good items are unambiguous and they

- refer to specific behaviours,

- avoid as far as possible terms of frequency or feelings and yet

- evoke instantaneous response (Kline, 1995).

7.5.6 Policy Recommendations for Personality Questionnaires

Validity, Reliability: As mentioned earlier personality questionnaires are not renowned for their high validity. In particular, projective tests are known for their low reliability and validity evidences.

Therefore, it is not recommended that personality questionnaires or projective tests be used as the sole selection method for *ab initio* trainee controllers.

Guideline 30: Personality Testing

It is not recommended that personality questionnaires or projective tests be used as the sole selection method for *ab initio* trainee controllers.

Due to some serious disadvantages of personality inventories, care must be taken in order to correctly interpret personality data during the decision on selecting *ab initio* trainee controllers.

7.6 Electroencephalogram (EEG) Measurement

EEG refers to the measurement of the overall state of the brain, i.e. of its continually changing responsiveness to input, including the interaction of that input with existing perceptual, cognitive and memory states (Heron, 1990).

EEG is not a test, it is a physiological measurement. Recent trials have been undertaken to make use of EEG measures for selection of *ab initio* trainee controller purposes (Weinberg, 1995; Heron et al., 1995).

However, this is still at the research stage and the study is still in progress. With regard to the validation of these measurements, this STF was unable to find enough empirical evidence of their validity to support their use. The STF concluded that extensive psychometric effort is still needed in investigating EEG measurements before they can be utilised in the selection of *ab initio* trainee controllers. Future progress in this field may require a review of the subject.

Guideline 31: EEG Measurement

At present EEG measurements cannot be recommended for selection of *ab initio* trainee controller purposes.

7.7 Pre-training Screening Tests

Several countries have tried to follow a completely different approach in controller selection, called pre-training screen (e.g. US, Netherlands) where candidates, after being pre-selected on the basis of pre-training screening tests, are hired into training for a number of weeks. The final selection is done after this pre-training screen.

The STF decided to concentrate on the other approach, the direct selection of candidates for *ab initio* training on the basis of psychometric tests.

The main reasons for this decision are the following:

- pre-training screening is not widely in use in Europe and elsewhere and hence the experience with this approach is still limited;
- current results of studies done on this approach are not as straightforward and consistent as one would prefer to arrive at a clear decision as to the general suitability of this approach;
- pre-training screening has a number of undesirable features for the candidates (e.g. leaving current job or study for a longer than usual period and other social disadvantages);

- current failure rates in the pre-training period and in later stages of the 'real' training are still rather high;

- it is as yet unclear whether instructors / trainers can be trained and can act as efficient predictors of success in later training stages.

Candidates for pre-training screening are selected on the basis of some sort of psychometric test(s) aiming at prediction of success in the pre-training screening phase. However, it is questioned whether psychometric tests with proven predictive validity for conventional ATCO's success in training would not be more efficient in selecting candidates.

The STF acknowledges the work done in this area and fully supports any studies which aim to clarify the issue further.

Page intentionally left blank

8. SELECTION INTERVIEW

The STF recognised that interviews are the most widely used and appreciated method for selecting candidates for *ab initio* controller training both by interviewers and interviewees. It is this source of data, together with those from standardised tests, that are used to gather information and impressions used in the later stage of decision-making.

The interview is also used to provide information to candidates.

The STF is of the opinion that there is no final solution for the selection interview. Different approaches to interviewing can in principle be followed. It is also stressed that interviews are in general a broad category of diagnostic instruments, which are applied in the form of a specific dialogue or interaction (question and answer).

Depending on

- the constructs that an interview aims to tap,
- the functions it is supposed to fulfil,
- the methods and modes applied in interviewing,

there is a great variety of interview forms.

This Chapter outlines the assumptions and principles in the development of best practice in selection interviews and gives practical recommendations for its application.

8.1 General Overview

Although the interview method is and will remain of great importance in selection, data on reliability and validity of this method are usually disappointing. The following overview addresses some of the main reasons for this problem and specifies principles of good practice in the choice, development, and use of interviews.

8.1.1 Content Validity of the Interview

The principles to be followed in the development and choice of interview are the same as in other selection procedures.

As other selection methods, the interview should support the inference with regard to the suitability of an applicant for controller training and job. The use of job information (e.g. from job analysis or CIT; see also Sub-chapters 8.2.4 and 9.5.1) as a basis in interview development is the best approach to assure the content validity of the interview.

Guideline 32: Content Validity of the Interview

To increase the content validity of the interview the interview should be training- and/or job-related.

8.1.2 Scope and Content of the Interview

An interview - as other selection means - is not a general method to accurately assess all relevant qualities of applicants nor is an interview superior to other, more standardised and valid sources of information on applicants as, for example, selection tests.

The content of the interview should be restricted to those characteristics of applicants that

- could not be assessed more effectively by other methods (e.g. by a test or by an application form);
- can only be assessed in an interview situation (e.g. require an interaction between interviewers and interviewee);
- are empirically related to success in subsequent training and/or employment (predictive validity).

The last bullet refers to the principle that empirical evidence should be sought to justify the use of the interview as a selection procedure.

Two areas of applicant characteristics that are relevant and can be assessed by interviewing applicants are:

- applicant characteristics that are related to training success (e.g. behaviour- and motivation-related to learning and examination, educational achievements, vigilance, communication and teamwork skills);
- criteria-related to training and job tenure (e.g. job and career motivation, general motivation, stress tolerance, decisiveness).

These characteristics are of high importance in *ab initio* trainee selection due to the high costs involved in training.

Guideline 33: Scope and Content of the Interview

Only those characteristics of applicants which cannot better be assessed otherwise and which are relevant for subsequent success in training and in employment should be assessed by interviewing. The use of the interview procedure should be validated.

8.1.3 Role of the Interviewer

Interviews are supposed to assess certain qualities of candidates which are relevant for subsequent training success and employment. Interviews are not meant to directly predict these future achievements from the answers given by applicants as a response to a certain question.

This rationale is particularly worthy of mention since some interviewers believe that they could directly "see" from an interview whether an applicant "would make a good controller or not". There is in fact no evidence for this belief.

Interviewers should therefore act as a kind of instrument to assess certain characteristics of applicants. It has been found in research that effective interviewers closely follow prescribed rules for the interview.

The following general principles apply with regard to the function and ability of interviewers:

1. Interviewers need to clearly and precisely know what characteristics or qualities of applicants should be assessed and how to assess them. Interviewers need therefore regular and specific interviewer training.

2. Interviewers can assess certain qualities of applicants better (i.e. more reliably) than others. Interviewers can, for example, assess English language skills more reliably than social skills of applicants. This should be taken into consideration in the design of the interview.

3. Interviewers are more able to assess certain qualities of candidates than to make a prediction with regard to the future success of the applicants. For example, they can assess social skills and motivation of applicants but they cannot directly predict whether applicants will be able to separate aircraft by using radar.

Guideline 34: Role of the Interviewer

The task of interviewers should be clear and should be described as precisely as possible. Interviewers should be trained on a regular basis.

The interviewer's task should be to obtain reliable assessments on certain characteristics of applicants. The characteristics should as far as possible be observable or otherwise easy to assess.

Interviewers should not be requested to directly predict the future success of applicants.

8.1.4 Increasing the Reliability and Validity of Interviews

Since reliability places an upper limit on validity, interviews that are more reliable will help to achieve a better prediction of suitability of applicants. Thus, higher validity of the interview may be achieved by

- increasing the reliability of the interviewer ratings,
- increasing training / job relatedness of the interview.

Concerning the first bullet available data clearly shows that structured interviews are more reliable. They give a more consistent account of the interviewee and result also in a higher interrater reliability. Structured interviews are also considerably more valid (predictive validity) (see the following sub-chapter).

The second bullet indicates that the interview should be based on training and job analytic information as mentioned before.

8.1.5 Interview Structure

Structured interviews are based on relevant job information from TA and aim at higher objectivity and reliability by minimising the chance for irrelevant information to be gathered and for bias or prejudice to occur which might otherwise influence the assessment. This will, as mentioned in the previous Sub-chapter, increase the validity of the interview as well.

The interview can be structured in two respects:

- by structuring the interview: using a rather standardised interview form (e.g. asking only predetermined questions in the same, prescribed way);
- by structuring the interviewer assessment: using rating scales or categories to work from and using trained interviewers.

This Chapter gives details of two different forms of structured interviews:

- the Situational Interview (SI),
- the Biographical-oriented Interview (BOI).

The possible combination of different forms of interview and the combination of interview with Assessment Centre (AC) are considered in Sub-chapter 9.8.

Completely unstructured interview forms are not considered in this Chapter although it is recognised that at least part(s) of the interview will in nearly all instances be informal and therefore unstructured or not standardised (e.g. “warming up” of candidates at the beginning of an interview; see Sub-chapter 9.8.1).

Guideline 35: Interview Structure

From a practical point of view, a structured interview should be used as far as this is possible.

8.1.6 Interview Panels

Board or panel interviews are slightly more reliable and hence their use increases the validity compared to individual (one-to-one) interviews. However, evidence from research has shown this effect is relatively small in case of well-structured interviews. In the case of structured interviews, one-to-one interview is often the practice.

The interview panel can include trained ATCOs, psychologists or both, or other trained personnel.

8.1.7 Separation of Interview from Other Data Sources

The selection interview is in many instances part of a decision-making process at the end of the whole selection process.

For example, in a multi-stage selection process with the interview as the last stage, it will very often be the case that interviewers are also involved in the final selection decision after the interview, where all information (e.g. test results, interview results) from applicants is considered. In this case the interview can be kept as an additional and independent source of information.

As a general principle, it would be preferable that applicant data from earlier stages in the selection process are not considered by interviewers **before** they have made their assessment of the interview performance of applicants (see also Sub-chapter 7.1 in the TS).

This principle is based on strong evidence from research. If interviewers have, for example, test information from applicants available before the interview, their assessment of candidates after interview is not better and frequently less valid than if it was based on interview information alone.

A different case exists where interviewers are as well involved in the pre-selection for the interview based on test results and/or have all applicant results from preceding stages available during the interview which means that the interview is not kept as a complete independent source of information.

In these cases, it will not be possible to establish clear and unbiased evidence on the additional (incremental) validity and utility of the selection interview due to statistical reasons.

Guideline 36: Separation of Interview from Other Data Sources

In order to comply with statistical requirements in establishing the unbiased incremental validity of the selection interview, the interview should be kept as an independent and additional source of data on applicants. Interviewers should therefore preferably not have detailed information from tests and other assessments before the interview.

8.2 Situational Interview

The SI is a rather recent method for the selection of personnel for a variety of jobs and has in recent years been considered as a possibility in controller selection.

This Sub-chapter addresses the underlying notions of SI and the considerations that should guide practical application of this tool.

8.2.1 Guiding Principles

The SI is a specific type of a structured and standardised interview based on hypothetical situations. It is a standardised instrument because all interviewees are presented with the same situations to which they have to respond. In addition, they are asked the same questions in the same way and order. The answers are scored following prescribed rules and procedures.

8.2.2 Objectives

The main objectives in developing and using SI are to

- increase interview objectivity and reliability,
- increase the predictive validity (power) of interviews.

In order to achieve a higher interview objectivity and reliability, the aim is to develop a structured and standardised form of the interview, to train and calibrate interviewers and to apply highly standardised scoring procedures.

The specificity of the SI compared to other approaches in interviewing can also be seen from the fact that it is a quantitative method: it follows the same principles of test construction (e.g. item analysis, internal consistency of items) as other quantitative measures of selection (e.g. psychometric tests).

With regard to predictive validity, the aim in developing SI is to derive situations (e.g. by a job or task analysis, by asking Subject Matter Experts [SMEs]) which correspond to a high degree with a given training or job reality. These hypothetical situations are then later used to provoke realistic behavioural examples from applicants during the interview.

Expressed behaviour intention (applicant response to hypothetical situation) is assessed in order to identify the degree of adequacy of the intention in regard to a similar training and/or job situation.

8.2.3 Basic Assumptions

The SI is based on Locke's (1968) theory of goal setting. The theory postulates that behavioural *intentions* are related to actual behaviour. It is further assumed, that intentions precede actual behaviour.

It is therefore assumed that how a person would act in a certain situation can be partly inferred from his/her expressed intention how he/she *would* actually act in this or in a similar situation or how they would tackle a problem- or task-related to the task or job.

8.2.4 Principles of Development

The first aim should be to identify training and/or job situations or incidents which are relevant and critical in terms of importance for success in training and/or in the job and which can be answered without previous ATC knowledge.

Critical Incident Technique

A technique frequently used for this purpose is the CIT (see Annex "Glossary of Terms"). It requires those knowledgeable about the ATC job and preceding training to describe the critical job or training incidents that represent effective or ineffective performance or behaviour. These incidents are then used to develop relevant situational questions for selection purposes.

The CIT often solicits performance extremes (e.g. very good or very bad performance) and omits examples of average performance. However, situations which normally occur very rarely might not necessarily be appropriate to identify "typical" training or job situations.

Applicability in Selection

It is also necessary to ensure that the situations can be used in selection. It has to be taken into account that applicants will generally have limited knowledge as to what the situation in controller training and on the job actually is. It is therefore important to reassure applicants that they do not need special job experience or knowledge to respond to the situations. Otherwise the situation will appear artificial to the applicant. The responses given might be totally irrelevant in this case and will fail to provide valid information that can be used in a selection decision (see also Sub-chapter 2.3.5.5 in EATCHIP, 1997).

It is further necessary to ensure that applicants are able to "picture" the described situations. Abstract situations will provoke arbitrary and irrelevant responses.

Assessment Scales

The answer of the applicant is normally compared to a number of normative answers that have been obtained from a group of job experts who have been asked to describe how 'good' and 'bad' trainees or job incumbents would act in each situation.

A range of possible reactions to the situations (which will cover answers that 'normally' can be expected) will be necessary to allow for an easy and correct classification of answers. It is at least necessary to provide interviewers with anchors for the scales for both "good" and "bad" answers for each situation.

The variety of possible answers will depend to a certain extent on the type of situations given (e.g. social interaction problems, interpersonal conflicts, job problems).

Item Difficulty and Variance

The difficulty of the items needs to be controlled. For example, items which lead to stereotype answers (e.g. nearly all applicants answer in the same way), show bottom effects (nearly all applicants receive a very low score) or ceiling effects (nearly all applicants receive a very high score) should be excluded or redesigned (see also Sub-chapter 3.2 of the TS document).

Social Desirability

It must be kept in mind also, that - as in other selection interview situations - responses may be biased due to the selection situation and that candidates could tend to produce responses they think are favoured by interviewers (social desirability).

As a general principle, it might be wise to include items in SI that can be used to assess possible distortion of answers due to social desirability and to train interviewers to verify this tendency in the answers of applicants.

8.2.5 Practical Application

The interview is normally led by one specially trained interviewer. She/he may be assisted by another trained person who concentrates on the answers given by the candidate. Only one candidate is interviewed at a time.

Situations are read to the applicant. After that applicants are asked to give a verbal expression of their reaction. The standardised question is addressing the type of response required (e.g. what they would do in this particular situation, how they would respond, how they would tackle the problem).

The typical task of the interviewer is to compare the applicant's answers with a given range of normative answers derived from training or job experts and determine which category comes closest to the answer given.

Instead of categorising applicant responses or in addition to this, interviewers can also directly score the adequacy of the response on rating scales.

8.2.6 Policy Recommendations

Reliability and Objectivity: One of the strengths of the SI is that all applicants receive identical questions and the information obtained from applicants is evaluated in a structured way, assuring a high degree of objectivity. These are necessary preconditions for an interview method to be reliable and valid.

Face Validity: Provided that the situations in SI are based on clearly identified training and/or job situations, the SI will be viewed as having high face validity both for those applying it and for those being interviewed.

Acceptability: However, it has been shown that applicants if they could choose between an unstructured interview and structured or situational interview, would prefer the unstructured interview form. This is due to the fact that interviewers using the unstructured form are seen as easier to influence and that applicants feel more comfortable with interviewers.

Predictive Validity: In some studies in Sweden, the SI has shown a discriminative validity, i.e. it discriminates between ATCOs and applicants. It was possible to correctly identify more than 80% of a sample of qualified ATCOs and applicants for ATCO training (Brehmer, 1996). The results of the predictive validity are not yet known.

However, if the SI is chosen as a selection tool, it should be kept in mind, that it is not the form of the interview which assures its predictive power, but the relevance and importance of the situations for relevant success criteria to be predicted, such as performance in training and/or in the job.

It also has to be realised, that, as with other forms of interview, a good structured and carefully designed SI does not prevent the poor administration and application (e.g. by unsuitable or not sufficiently trained interviewers, see below).

As research on SI has shown, a wrongly developed SI can even have an inverse effect in terms of validity if irrelevant job situations are used.

Practicality: When it comes to applying an interview method, which is psychometrically and legally defensible as well as being practical in application, the SI has shown to be an efficient interview method.

Combination with other Methods: The SI can be used in combination with a biographical or structured interview but as well as the sole interviewing method.

Interviewer Selection and Training: As with other approaches in interviewing applicants, it must be ensured that the SI is administered only by personnel which has been selected on the basis of their interviewing skills and have

been appropriately trained in the use of the SI. Regular calibration training should be done to reduce well-known interviewer effects (e.g. halo effects) and inconsistency in the application of SI.

Assurance of Validity and Reliability: As other selection tools, the SI needs to be followed up and controlled in regard to its effectiveness and efficiency. This includes separately the interview (e.g. validity of situations and/or answer categories) and the interviewers. Poor interviewers should be identified and either retrained or excluded.

Guideline 37: Situational Interview

If it is intended to implement the SI as an interviewing method, it must be ensured that situations are used which are applicable in the selection situation and are valid for success in training and/or job proficiency.

It must be ensured also that the answers given allow for a reliable assessment of behavioural intentions that are specific for the task at hand and are valid for the success or failure in the situation described.

8.3 Biographical-oriented Interview

The Biographical-oriented Interview (BOI) is a traditional interview method used in employment. Older versions of this interview form focused on generalities about qualifications of applicants to get an impression of their competence but were not predictive of success in a specific job. Another feature of this interview form was that it was unstructured.

However, the BOI has been further developed into more structured and standardised forms with the focus on those behavioural characteristics that are predictive for training and/or job success.

This Sub-chapter addresses the underlying notions of the structured or completely standardised BOI and the considerations that should guide practical application of this tool.

8.3.1 Guiding Principles

The guiding principle in the development of BOI is to construct a method, which allows assessment of past behaviour of applicants as reported to interviewers with regard to certain performance or behaviour criteria.

It is further assumed that on the basis of this assessment, a prediction can be made towards the tendency of an applicant to behave in a similar way in comparable situations in the future. Unfavourable behaviour tendencies may create risks in training and/or the job.

8.3.2 Objectives

The BOI aims to make predictions on the future behaviour of applicants in important situations based on assessments of the behaviour applicants have shown in the past.

8.3.3 Basic Assumptions

The BOI is based on the assumption that information from persons about past events (e.g. their past achievements, success or behaviour) allows prediction of events in the future (e.g. their future achievements, success or behaviour).

The basis for this assumption is the theory of behaviour consistency, which states that individuals behave in a consistent way in similar situations.

It is therefore assumed that if persons are put in the future into a certain situation (e.g. into a learning situation in training) they would react or behave consistently in a similar way as they did in the past in a comparable situation (e.g. in school).

8.3.4 Principles of Development

The main objectives in developing and using a BOI are to

- identify relevant behaviour domains in the past that are good predictors for future behaviour (to ensure predictive validity);
- make sure that this behaviour domain chosen “overlaps” with a certain relevant criterion domain (to ensure content validity);
- identify relevant behavioural criteria against which the behaviour reported by applicants can be assessed (to ensure criterion-related validity);
- make sure that the behaviour in question can be reliably and accurately assessed by interviewers with reference to the criteria identified (to ensure consistent scoring).

Relevant Behaviour Domains

The following list gives some examples for behaviour domains, which might be considered for developing and using a BOI for *ab initio* trainee controllers.

<u>Performance / Achievement Behaviour</u>

- Performance and achievements in school, apprenticeship, university, etc.
- Working behaviour (e.g. preparation of exams, risk taking behaviour)
- Achievements in leisure time activities (e.g. sport)
- Performance / achievements in past jobs

Vocational Motivation / Interests

- Length and stability of job interest (e.g. history in past jobs, instability and inconsistency in study subjects)
- Knowledge about the job
- Realistic job view and career expectation
- Dealing with vocational key situations

Leadership / Planning / Decision-making Behaviour

- Initiative in school, leisure time activities, etc.
- Experience in planning and organising activities
- Experience with leadership roles in school, university, extra-curricular activities, etc.
- Decision-making in school, vocation, etc.

Communication Behaviour and Teamwork Skills

- Communication behaviour in groups, family, job, etc.
- Experience in mixing with others (building relationship)
- Experience with teamwork

Stress Handling

- Coping / dealing with stress in school, university (e.g. exams) and home
- Coping / dealing with stress in leisure time activities (e.g. sports)

Validity

The development and use of the BOI in selection needs can first be justified on the basis of a relationship between the content of the biographical event or personal history experience and the content of the training and/or the job. It has to be shown that certain achievements, experiences or prior training (e.g. in school) are related to the training and/or job for which they are evaluated or required (SIOP, 1987).

Empirical evidence on the validity of the BOI can and should be sought and ensured by collecting and analysing interview data on a single item level.

Available evidence from research shows that the BOI can reach a sufficiently high criterion-related validity. The validity is less high with younger applicants who have not yet established stable behavioural patterns.

Objectivity and Reliability of BOI

To achieve a higher interview objectivity and reliability of the BOI, the aim should be

- to develop a structured and standardised form of the interview;
- to apply standardised scoring procedures;
- to separate information gathering and assessment (during interview) from decision-making (after the interview);
- to train and calibrate interviewers for interrater reliability, aiming at a high level of reliability (e.g. 0.85).

The structured and standardised form of this interview would require that all applicants be asked the same questions in the same order. This approach ensures that interviewers collect the same information from all applicants and avoids that snap and possibly erroneous judgements are made.

Any decision as to the suitability of applicants should be taken after the interview has been conducted. The information gathered and the assessments made should be considered, interpreted and if necessary adjusted after the interview by the interview panel with the aim of arriving at a balanced and justified decision on applicants.

8.3.5 Semi-structured Interview

As a compromise between unstructured and fully structured interviews, a semi-structured interview form could also be developed. The semi-structured interview uses interview questions, which are prepared in advance and the same questions are asked of all applicants. However, follow-up questions are allowed to probe specific areas in more depth.

With regard to scoring procedures, interviewers should be given specific criteria and a scoring key to evaluate responses to each question (e.g. on a behavioural-anchored rating scale). This approach assures that all applicants are evaluated against the same criteria and avoids that irrelevant criteria are used which minimises biases (e.g. the halo effect).

8.3.6 Questioning Techniques

To gain reliable and selective information of applicants a certain questioning technique should be followed. The following technique is geared towards gaining behavioural information about applicants:

Concrete examples of past behaviour can be discovered following a triad structure explained in the following (after Ghiselli, 1966):

1. **Address initial situation**

   Example: Address school situation and performance in school; ask applicants what their achievement positions were in general compared to others in their class in (lowest third, medium, upper third); ask which subject was the most difficult one for them; ask about exams in their last two years of school in this subject; ask for their ambitions in this subject.

2. **Address actions taken**

   Example: Ask how applicants prepared for exams in the last two years in the subject with the most difficulties.

3. **Address outcomes of actions taken**

   Example: Ask how applicants performed in exams; whether they achieved what they wanted to achieve.

8.3.7 Interviewer Training

Interviewers need to be trained in

- what is their role and task,
- how to conduct the interview (the interview process),
- what they should look for (the criteria),
- how they should look for it (the methodology applied),
- what are common sources of bias and possible counter measures,
- how to use ratings scales efficiently,
- how to apply appropriate questioning techniques and style of questioning.

Active training should be given to interviewers by including video feedback, role-play and conducting first interviews under supervision.

It is important that interviewers through their reaction should only stimulate applicants to further exploration. **Any reaction of disapproval or approval to applicant's answer should be avoided**.

An important aim of interview training should concentrate on the role and task of interviewers. Their main task should be to collect and assess information.

8.3.8 Practical Application

The BOI should be conducted by a specially trained interviewer. It is preferable that the interview is conducted by a panel of trained interviewers. Panel interviews produce slightly more consistent results.

Interview Process

Applicants are asked about a certain aspect of their life (e.g. how they prepared for exams in the last two classes in their school career) and requested to give a description of their behaviour.

If possible more than one performance area should be used to establish the behaviour tendency of an applicant (e.g. preparation for exams in school, preparation for exams in professional training as, for example, in apprenticeship or university courses).

The task of the interviewer is to assess the behaviour with regard to relevant dimensions (e.g. tenacity, flexibility, solving upcoming problems) on ratings scales (e.g. very low, low, moderate, high, very high).

Scoring Procedures

The rating scale can be anchored as shown in the following example:

Example:

Question

"What did you do to prepare yourself for exams in the last two classes of school?"

Answer Rating

1) In preparing for exams the **efforts** spent were:

1	2	3	4	5	6
No efforts	I	Just sufficient	I	High	I
	Very little efforts		More than sufficient		Very high

2) In preparing for exams the applicant **solved upcoming problems**

1	2	3	4	5	6
Not at all	I	Less than adequate	I	More than adequate	I
	Inadequately		Adequately		Perfectly

The behaviour can also directly be scored with regard to the adequacy of the behaviour towards a goal (e.g. to pass exams, to achieve a good grade), using rating scales.

The scores from several dimensions might be integrated after the interview into an overall assessment of risks regarding training success, job tenure, etc.

8.3.9 Policy Recommendations

Reliability and Objectivity: Provided a consistent approach is used (e.g. use of structured form, identical questions, assessment done in a structured way), a sufficiently high degree of objectivity and reliability can be achieved.

Face Validity: Provided that the questions in BOI are based on applicant characteristics, which are perceived as being related to training and/or job performance, the BOI will be viewed as having face validity both for those applying it and for those being interviewed.

Acceptability: According to some findings, most applicants feel comfortable with BOI. There are some indications that interviewees feel that strengths and weaknesses of applicants can be correctly assessed by this interview form.

Predictive Validity: It has been found that BOI can reach a sufficient high level of predictive validity provided the interview questions used for making assessments have shown to be relevant for success in training and/or in the job.

As with other forms of interview, a well-structured and carefully designed BOI does not prevent poor administration and application (e.g. by unsuitable or not sufficiently trained interviewers).

Practicality: When it comes to applying an interview method, which can be designed as being psychometrically and legally defensible as well as practical in application, the BOI has shown itself to be an efficient interview method.

The costs of developing BOI are reasonably low.

Combination with other Methods: The BOI can be used in combination with other interview forms. In most instances, it is currently used as the sole interviewing method.

Interviewer Selection and Training: The BOI should be administered only by personnel who have been appropriately trained and have reached the required level of proficiency.

Regular calibration training should be done to reduce well-known interviewer effects (e.g. halo effects) and inconsistency in the application and in the interpretation of information.

Assurance of Validity and Reliability: The BOI needs to be followed up and controlled in regard to its effectiveness and efficiency. This includes both the interview (e.g. validity of interview questions and/or assessment ratings) and the interviewers. Poor interviewers should be identified and either retrained or excluded. Interrater reliability should be high (e.g. close to or above .85).

Biased Answers: There is evidence that the tendency of applicants to report behaviours they think are favoured by interviewers is reduced in BOI due to the fact that questions

- address concrete events within the biography of applicants;
- are based on real experience;
- could in principle be checked or investigated (e.g. against actual school grades).

Guideline 38: Biographical-oriented Interview

If it is intended to implement the BOI as an interviewing method, questions should be used which address significant sections within the biography of applicants where applicants have established consistent and stable behaviour.

Past behaviour reported by applicants should be valid for training and/or job behaviour. The interview should be structured or semi-structured and used only by trained interviewers. Assessments should be made using rating scales on relevant behavioural criteria. The process of collecting information (during the interview itself) and making decisions (after the interview) should be clearly separated.

Page intentionally left blank

9. ASSESSMENT CENTRE IN SELECTION

An Assessment Centre (AC) is a method to assess actual, observable behaviour of persons in simulated situations. The AC is a group-oriented and standardised series of activities organised in exercises known or believed to be relevant to the work performed in an organisation.

ACs typically involve participants, in groups of three-twelve persons, completing exercises which simulate activities that are carried out in a certain job. Participant behaviour is observed and assessed during these exercises by trained assessors.

In AC the focus is on assessment of behaviour rather than on behaviour intentions, compared to the SI, or on abstract characteristics (traits) of persons, as in some personality questionnaires or interview forms.

9.1 Background

AC originally was developed and used for the purpose of assessing certain personal characteristics in a selection situation. However, today most AC exercises have at least some development aspects to it.

The development aspect of AC becomes obvious as individuals are supposed to be involved actively and to participate in the AC process rather than being a passive recipient of it. They receive also extensive and concrete feedback after the AC has been conducted with regard to their performance in the process.

This tool is used in nine countries at least for the selection of *ab initio* trainee controllers (EATMP, 2000a).

Generally, and at a later stage in the career of controllers, the AC can be used to identify further potential of persons with a view to subsequent training and development and/or career advancement in other jobs (e.g. supervisor, managerial or administrative jobs, instructor job).

This Chapter addresses the underlying notions of AC and the considerations that should guide practical application of this tool if this is intended.

9.2 Guiding Principles

The guiding principle behind AC is that the best way of assessing the behaviours of a person which are relevant for job performance is to have applicants for the job carrying out tasks which sample the tasks required in the job and are as similar to them as possible.

9.3 Objectives

The objectives of AC will depend on the purpose for which the AC is designed and applied. It must be borne in mind that the AC can in principle be developed for different purposes (e.g. for selection purposes, for development purposes) and a multitude of objectives.

The general aim of the AC is to give a better picture of applicants by involving multiple assessment methods and multiple trained observers to assess current, observable behaviour. This can lead to better decisions on the suitability of applicants.

Thus, the AC aims at a higher validity in terms of higher accuracy in predicting future behaviour.

Other objectives are to familiarise candidates with the job, to improve their self-estimation and self-awareness, to address further training and development needs, and to give the opportunity for social comparisons.

9.4 Basic Assumptions

The basic assumption in AC is that the best way to observe and assess behaviour is to observe and assess it as directly as possible in relevant situations which are simulated.

It is further assumed that the behaviour in simulated situations in AC directly pictures the behaviour of participants they might show in 'real-life' situations. The AC approach acknowledges the fact that behaviour of individuals depends to a certain extent on the actual situation they are in and on other persons with whom they interact in this situation.

Based on these assumptions it becomes obvious that AC procedures need to be well designed for

- the purpose,
- the people to be assessed,
- the samples of situations to be simulated and
- the relevant behaviour used in 'real life' situations in the job.

9.5 Principles of Development

As with other methods in selection of trainees (see Sub-chapter 8.2 on SI) the AC needs to be specifically designed for the purpose of selecting *ab initio* trainees.

In doing this it has to be ensured that

- the AC exercises and the behaviour used in the exercises are relevant to the job of a controller;
- overt behaviour can actually be evoked in a simulated situation (the AC exercise);
- the exercises are applicable and can be observed by assessors;
- the exercises can reliably and consistently be assessed by trained assessors.

9.5.1 Relevant Behaviour Domains

The particular behaviours used in the AC tasks need to be clearly identified. Relevant behaviours should be determined through a task and job analysis or by CIT. The latter helps to identify critical behaviour dimensions of controller performance.

The requirements of AC as outlined above are not easily met and it must be borne in mind that considerable effort and resources are necessary to achieve them.

The following is a list of behaviours which are typically mentioned as assessable in AC:

<u>Leadership / Planning / Decision-making</u>

- Leadership behaviour / style
- Initiative
- Planning and organising activities / actions
- Problem solving
- Decision-making

<u>Communication / Cooperation / Teamwork</u>

- Adaptability and flexibility in groups
- Communication behaviour / style
- Working in a team
- Dealing with conflicts in teams
- Assertiveness

Stress Handling

- Coping with stress
- Dealing with time constraints
- Action planning

9.5.2 Development Criteria

While there are many forms of assessment procedures, the term "Assessment Centre" has a specific meaning. According to the 'Task Force on Assessment Centre Standards' (TFACS, 1980) the following criteria must be met in developing an AC:

1. It must contain an explicit definition of the relevant determinants of task performance. These are the behaviour dimensions which are to be assessed. The relevance of the behaviour dimensions should be ensured by a job analysis.

2. Multiple assessment techniques, including simulated tasks, should be used. These are the exercises which represent important aspects of the job in which the applicant is actively involved and which stress the identified behaviour dimensions.

3. The performance of applicant behaviour is observed and assessed by a selected and trained team of assessors who are knowledgeable about the tasks and the job. The team should be trained to observe, evaluate, report and feed back the behaviours.

4. The integration of all information occurs after all assessment techniques are completed, at an assessment evaluation session. In these sessions, the assessment ratings made on the behaviour dimensions are pooled and discussed by assessors before an overall assessment with regard to the suitability of the applicants is made.

5. Each dimension evaluated in AC can be measured by more than one method.

9.5.3 Exercise Examples

The following exercises are typically used in an AC:

- in-basket exercises (e.g. responding to various issues related to the job, finding solutions for problems);
- group discussions with or without assigned roles (e.g. discussing the requirements of the job);
- teamwork exercises (e.g. working as a team of task force members);

- individual presentations (e.g. presenting own ideas on certain job-related topics to an audience, responding to upcoming questions);

- role-play (e.g. acting as a supervisor in solving a conflict between team members);

- group exercise in action planning and problem solving (e.g. planning and preparing a work process and finding solutions for upcoming problems).

It should be borne in mind that an increasing number of applicants have received training in AC and thus know about this method and what are 'preferred' behaviour styles which would normally lead to good assessments. There is some evidence that 'pressured' exercises, that is, exercises that give limited time for preparation or follow each other in a series help to overcome these training effects.

These exercises are in principle suitable as a framework for designing ATC specific exercises which could simulate demands in

- Initial Training (e.g. working in a study group with other trainees);

- Unit Training (e.g. following instructions, dealing with instructor feedback);

- job demands (e.g. cooperating with others in a team, task sharing).

However, the number of exercises will necessarily be small and it can be questioned whether they sufficiently sample the training and/or job content. This must be regarded as a serious limitation of the AC as a selection procedure compared to the use of AC as a development tool.

9.6 Practical Application

9.6.1 Assessment Process

An AC usually involves several assessors who evaluate each candidate on each selection exercise.

The AC is group-oriented because

- the assessment of competencies is carried out in groups of candidates, that is, between three and twelve persons are assessed simultaneously;

- the assessment is carried out by groups of evaluators or assessors who work in teams to observe and evaluate applicants.

Applicant Instruction

Applicants are oriented as to the nature and purpose of the programme.

Observation and Initial Evaluation

Each assessor is assigned to a number of applicants he or she observes and will report on. No applicant should be exclusively observed and evaluated by a single assessor.

Each applicant is evaluated separately. Individual assessment reports are prepared immediately after each exercise.

Integration of Information

It is common practice in AC that the process of observing applicants and making initial assessments (done by each assessor) is kept separate from final evaluation and decision-making. The latter is carried out by the assessor team after all AC exercises have been conducted on all applicants.

Assessors pool their observations and assessments on each candidate across all exercises. Ratings are discussed and differences are resolved.

Finally, a common final evaluation is reached by discussion.

Feedback to Applicants

Normally, feedback is given to candidates after the AC in which the candidate is informed about his/her performance in all exercises and if possible and appropriate, about their potential for future development and/or for other tasks.

9.7 Policy Recommendations

9.7.1 General

Applicant Training: An increasing number of commercial institutions offer AC training to applicants with the aim to train applicants in a kind of 'role-playing' during standard AC exercises. That is, applicants are trained to show behaviour known to be preferred by assessors leading to good assessments.

To prevent this, exercises should be specially designed or conducted and/or assessors should be trained how to detect the tendencies of 'role-playing' or reacting stereotypically in a social desirable manner.

Face Validity: Provided that the AC exercises have a direct link to requirements in training and/or the controller job, the AC will be viewed as having face validity. The exercises need to reflect important requirements of the controller job to allow for assessments as to how individuals might perform in the job.

Acceptability: If the AC exercises are specifically designed for the purpose of selecting *ab initio* trainees, participants will feel to have been fairly assessed.

There is evidence that applicants perceive themselves as being successful and able to show their potential in AC and that they enjoy the exercises to a high extent. The AC is viewed rather positively both from those applying it and for those being assessed.

Practicality: The AC normally needs a specific set-up and a well-orchestrated preparation before it can be actually conducted. This includes in some instances specific material, facilities, specifically trained assessors and considerable efforts and resources (e.g. time, money).

The time needed for conducting an AC is considerably lengthy. From this point of view, the AC is not an easy, convenient and a quick to apply selection method.

With regard to psychometric and legal requirements the AC can be designed to be an efficient method.

Costs: The AC involves rather high costs in terms of money and time to be spent before it can actually be used in selection. In addition, the costs in administration (e.g. personnel involved, facilities used) and maintaining the selection tool are rather high as well. AC is by far more expensive than other selection tools mentioned before.

A decision has to be made on the basis of cost-benefit considerations whether the development and implementation of the AC method is worth these costs.

9.7.2 Training

Assessor Training: Assessors need to be specifically trained and instructed in the process and the organisation of AC, the administration dimensions for assessments, etc. Regular training sessions should be held before each campaign.

The fundamental training of assessors should include

- observation and assessment training (e.g. observational techniques, observation exercises, systematic improvement of the observation - assessment process, the order of events of the AC programme, handling of observation / assessment tools, ratings errors, identifying social desirability behaviour);

- feedback and calibration training (e.g. feedback on assessor rating scales, correlation between behavioural dimensions assessed, differences between assessors).

Assessors also need to be trained in giving feedback to applicants.

9.7.3 Psychometric Issues

Reliability and Objectivity: The high unreliability of observers in behaviour observation can be largely attributed to well-known rating errors such as first impression and the halo effect.

It has been established that the interrater reliability was sometimes insufficient with 0.50. A sufficiently high reliability of about .80 or higher can only be achieved by regular calibration training of assessors.

Predictive Validity: Across all job families, ACs have very high validity for predicting promotion, but only a moderate validity for predicting actual performance. An average correlation of 0.43 in predicting later supervisor performance ratings has been established (Hunter and Hunter, 1984).

Caution must therefore be given to the appropriate exercises and type of AC to be applied. A carefully designed validation process is essential.

Construct Validity: The lack of construct validity of the ratings on behavioural dimensions (that is the constructs) is the main weakness in AC. There is enough evidence that the ratings of assessors on the same behavioural dimension are specific to the exercises and do not generalise from exercise to exercise.

This emphasises the fact that applicant behaviour is to a high extent situation specific. The AC is therefore only to some extent able to illuminate stable behaviour tendencies, skills and abilities of applicants.

Considerable attention must be paid to various aspects in AC including parameters such as type of exercises, the abilities required in exercises, the number and the kind of other participants, previous knowledge required in ATC, etc.

Sampling Problem: As mentioned above, the dimensions used in the behavioural observations and evaluation are highly situation specific. As one countermeasure, one should carefully select the situations simulated and control and ensure they represent the demands of the ‘real life’ situation.

However, most ‘real life’ situations are rather complex, demand different abilities, etc. This could be overcome by a ‘sampling approach’ with a sufficient number of different situations. This possibility is actually limited in AC due to its nature.

Guideline 39: Assessment Centre

If AC is intended to be used for selection of *ab initio* trainees, relevant, observable, reliable exercises should be developed.

Assessors in AC should be regularly trained and calibrated.

Final evaluation of AC should be taken into account with the results from other selection tools to arrive at a decision on candidates for intake to training.

9.8 An Integrated Approach in Interviewing and Assessment

It is obvious that there are no absolutes in interviewing applicants since the options available have strengths and weaknesses and/or cover several aspects of behaviour, which are equally important.

The interview as it is regarded in this document is in fact a framework, which can and should be filled with multiple contents and can be conducted using various techniques.

The following summarises the basic elements from SI, BOI and from AC described in the preceding Chapter:

- SI provides applicants with samples of situations in relevant areas and asks them for their intentions to behave in these situations;
- BOI searches for those relevant situations in past history, which are 'functional' for current and future behaviour in similar situations;
- AC simulates relevant situations and observes how applicants actually behave.

Obviously all three techniques can give important information on applicants and ideally most interviewers would prefer if they could use all three.

This Sub-chapter outlines this possibility and gives examples of how the combined interview - assessment process might look.

9.8.1 Combined Interview

This interview is a combination of all three interview / assessment techniques and requires two interviewers as a minimum, interviewing one applicant at a time.

The duration of the interview will be between 45 - 60 minutes.

Interview Opening

The interview should start with an informal talk with the applicant which aims to "warm up" the situation and help to put the applicant at ease.

Nevertheless, the opening questions should relate to the application and selection situation.

No assessments are made in this phase.

Self-introduction of Applicant

This part is designed according to an AC exercise (self-presentation) where applicants are requested to give a concise, but precise description of their career up to now and their goals and expectations regarding their future career.

Applicants report freely to interviewers about their educational and professional (if any) and extra-curricular background and on their vocational choice and expectations.

Interviewers make assessments on rating scales. The behaviour dimensions could be: verbal expression, nervousness, realistic expectations concerning ATC job and career, and vocational planning behaviour.

Biographical Questions

Interviewers ask (semi) structured biographical questions as described in Sub-chapter 8.3.

Assessments are made on rating scales on behaviour dimensions as outlined in Sub-chapter 8.3.

Situational Questions

This part should be designed according to SI as described before in Sub-chapter 8.2. The situation samples should relate to the content of future training and controller job as identified by CIT.

Applicants are asked to respond to situations by indicating how they would react in the situations.

Assessments are made based on behavioural-anchored rating scales.

Interactive Part

This part of the interview could be designed according to an AC group exercise. The interviewers act as members of the group with which the applicant needs to interact in order to be most effective and arrive at a solution for a given problem.

The applicant receives written information about the problem to be solved within a certain time frame. The information given might be incomplete or otherwise not easy to understand.

Interviewers offer their help and their expertise in finding a solution. The board is instructed to offer their help only on the request of applicants.

Another example for an interactive part may include a role-play, which involves the applicant as an ATCO and one of the interviewers as a 'pilot'. The applicant can be instructed to insist on delaying the aircraft. The applicants know that the pilot will try to argue and their task is to insist on delay in an appropriate, assertive way.

This interactive exercise can measure the understanding of keeping accuracy and safety as a prime concern, keeping the delay to a minimum, authority of tone of voice, conciseness, answers to the point and assertiveness in communication. Reasonable answers need to be defined.

The behaviour of applicants is assessed on dimensions like: initiative, interaction with interviewers, sharing own knowledge with interviewers, requesting information from interviewers, tenacity in trying to find a solution, nervousness.

Interview Closing

The interview board offers the opportunity to applicants to ask questions.

9.8.2 Information Integration

Interviewers pool and discuss their ratings and arrive at a common assessment in all behaviour dimensions rated.

Aggregation of values across all behaviour dimensions is finally done using a summative score. Only if pre-established weights exist from a validation study should these be used in aggregating the values.

In a final step interviewers integrate the weighted interview score with the test scores to arrive at a final decision on the applicant.

9.8.3 Control and Validation of Interview

Interview data should be gathered on single item level for further item analysis and for validation purposes.

The intercorrelations between ratings on same / different behaviour dimensions should be checked in particular to identify halo effects and other biases.

Page intentionally left blank

10. PROFESSIONAL PRACTICES IN TEST APPLICATION AND EVALUATION

This Chapter gives consideration to several critical issues in the application and evaluation of selection tests.

There is no one way of selecting candidates, but there are several widely accepted, general rules in psychometrics. This Chapter mentions some of these.

Sub-chapter 5.2 of EATMP (2001b) referred already in detail to professional practice of test administration, recommendations regarding the testing environment, test evaluation, scoring and reporting procedures and these are not repeated here.

Chapter 1 of EATCHIP (1997) described a professional method of evaluating tests or test batteries, which was adopted by the STF. This Chapter is recommended as a general outline in evaluating and considering tests or test batteries.

10.1 Guiding Principles in Setting Cut-off Scores

Setting cut-off scores is a matter of professional judgement rather than a scientifically derived technique. It will include considerations and estimations of cost-benefit ratios, number of openings and selection ratio, success ratio, social policies of the organisation, or judgement as to required or assessed KASOs on the job (SIOP, 1987).

The issues of reliability, validity and utility refer to the **distributions** of scores, the inferences drawn from these scores, and the accuracy of decision made, and not to specific **scores** used for selection purposes for example. From a legal point of view, the test user will have to show at least that the test is valid. In some countries the cut-off score used has in addition to be justified.

It should be kept in mind that it cannot be proven that an individual who scores slightly above the cut-off score is a better candidate than an individual who scores slightly below the cut-off score. However, if test scores and job performance are related (linearly), even small differences between scores have different probability in terms of expected job performance. If this can be shown (for example on the basis of an expectancy table), this information can be used to define cut-off scores.

These arguments strongly support the view that only tests and other methods with sufficient validity should be used for selecting candidates.

10.2 Practical Applications of Setting Cut-off Scores

Beyond legal and psychometric aspects, technical and practical problems associated with setting cut-off scores are a main concern. The following two different approaches (which can be used) are chosen for their simplicity, the minimisation of subjective judgements and the applicability in a multi-stage selection process. Screening at each stage of the process is felt necessary due to high costs or resources.

10.2.1 Human Resources Planning Approach

The projected personnel needs and past history of the proportions of applications accepted are considered in setting cut-off scores on tests to yield the number of trainees needed.

Cut-off scores are set at the point that provides a sufficient number of candidates in regard to recruitment needs and minimises adverse impact.

However, this method will almost certainly result in varying cut-off scores in each selection campaign, depending on the number of applications, the 'quality' of applicants, the number of candidates needed, etc. Trainee groups will be inhomogeneous.

It is not recommended to misuse the cut-off score just to obtain more trainees.

10.2.2 Norm-referenced Methods

Cut-off scores are set on the basis of the distribution of test scores (e.g. at the mean, one standard deviation above the mean, etc.) in a pre-defined norm sample.

More homogenous trainee groups are ensured with this method, where cut-off scores are applied for longer periods. However, the number of applicants for the next selection stage and for final selection will vary.

10.2.3 Criterion-referenced Methods

The general disadvantage of both the above-mentioned methods is that the relation between cut-off score and training and/or job performance is unknown or - if available - not used.

Both methods are therefore not acceptable if a minimum competency and proficiency is the objective in selection. It could be argued that this is the case in ATC where licensing and certification requirements have to be met by trainees later in their career.

In fact, by setting a cut-off score on tests, one tries to establish a standard which has to be achieved by a person in a test and which ensures that the person later achieves the standards in training and in the job.

Therefore, the objective in decision-making is to match a criterion and to ensure the smallest number of decision errors. The cut-off score on a test is selected to maximise the validity of a selection decision.

Figure 2 represents the two kinds of correct decision and the two kinds of incorrect decision that can be made in any selection situation.

	Persons who are later actually **unsuccessful** (e.g. failure in training)	Persons who later are actually **successful** (e.g. pass the training)
Persons **above** cut-off score in a test (positive candidates)	*FALSE POSITIVES*	***VALID POSITIVES***
Persons **below** cut-off score in a test (negative candidates)	***VALID NEGATIVES***	*FALSE NEGATIVES*

Figure 2: Correct and incorrect selection decisions

The following methods take these requirements into account. They presume that a standard already exists on an external criterion (e.g. criteria in training and/or job) and that this can be retranslated into a cut-off score on the tests.

Criterion-referenced Validity Methods

Use of Direct Performance Information: Certain criterion-referenced tests can be developed to provide direct performance information on certain skills or knowledge. Examples are English language or work sample tests. If this information can be derived, the cut-off score is chosen so that candidates having the required knowledge or skill are selected.

If this direct evidence of criterion performance from criterion-referenced tests is not available the following method can be applied:

A cut-off score is set, which delivers candidates with acceptable high **expected** training / job performance or success rate. This has to be evaluated.

Direct Use of Validity Information: The information on criterion-related validity can also be used to determine cut-off score directly from the minimum acceptable training or job performance level as a function of the **validity** of the test. This method can be described as follows:

- The regression of test scores onto training or job performance criterion provides the prediction equation for calculating the test score which is associated with the lowest acceptable performance criterion value or the highest acceptable failure rate (e.g. 10%).

This method is fully in line with the ‘weighted composite score’ approach, which is discussed in more detail in Sub-chapter 11.3 of this document and Sub-chapter 7.2 of the TS. The method can only be applied if complete and relevant validity information is available which takes several years to obtain.

However, this method is recommended to be applied in order to properly adjust existing pre-defined cut-off scores when this information is available.

Use of Expectancy Tables: The following method can be applied if complete data on success of trainees is available in an expectancy table.

The cut-off score is set at that point in the distribution of test scores that splits the sample into applicants with high success and those with low success (see Table 1).

Table 1: Determination of cut-off scores from expectancy table

TEST SCORE (Raw Score)	CRITERION		SUCCESS
	PASS	FAILURE	RATE
53	1	0	100%
52	3	0	100%
51	6	2	75%
50	7	3	70%
49	9	4	69%
48	12	6	67%
47	11	6	65%
46	10	7	59%
NEW CUT-OFF SCORE = 45			Mean=68%
45	7	9	44%
44	4	12	25%
43	2	4	33%
42	1	5	17%
41	0	3	0%
40	0	2	0%
39	0	1	0%
OLD CUT-OFF SCORE = 38			
38	..	..	Success Rate
..	..	..	with old cut-off
..	..	..	score was
15	..	..	53%

Note: The overall performance (or success rate) in this example could be improved by 15% (68% instead of 53%) with the new cut-off score; here the cut-off score is set at the point where success rate is below 50%. The new cut-off score rejects 36% of candidates compared to the old cut-off score. It is assumed that the test is part of a test battery. Therefore, it must be assumed that a further portion of candidates passing the new cut-off score is rejected based on other tests of the battery.

The method can also be applied if assessment information about trainees with regard to expected success or failure is available. This could be derived, for example, from instructors who classify trainees as “going to succeed / going to fail” before trainees are actually (finally) examined.

The proposed method again can only be applied if trainee data regarding success criteria is available. Again, the method is recommended to be applied also in adjusting existing cut-off scores.

It should be mentioned here that the figures obtained are not always as straightforward as shown in Table 1 above where the cut-off point can be clearly identified. However, if a test fails to discriminate between successful and unsuccessful candidates with regard to an important criterion, this must be seen as an indication for a lack of discriminant validity of the test and should lead to further investigations (for an example of discriminant validity, please see Sub-chapter 8.2.6 above).

Using expectancy tables to identify cut-off scores can also help to identify possible risks (e.g. higher failure rates) related to any changes in cut-off scores in order to select more candidates into the training system due to trainee shortages (see also Sub-chapter 11.4).

Criterion-referenced Cut-off Scores can also be identified with the following methods:

- A group of young, qualified controllers is tested before the test is used for selection of new candidates. A (new) cut-off score is identified in the distribution of scores from candidates that ensures a sufficiently high performance compared to controllers. However, the strong group identification (which occurs during training) has to be considered.

  It should be mentioned that this method is not always applicable to all controller age groups. It might for example be inappropriate to test more senior ATCOs with a sophisticated English language test, which is designed for school leavers. It will be inappropriate to use more senior controllers to identify cut-off scores in some memory tests.

A somewhat different strategy is used in the following example:

- A group of trainees is tested and are split by their instructors into two groups: those with insufficient training performance (e.g. likely to fail the course, marginal performance) and those with sufficient to high performance (e.g. likely to pass course, above minimum performance). Cut-off scores are set according to the method described in Table 1 above.

Due to specific skills and attitudes / personalities developed in training, the use of trained controllers in setting cut-off scores in tests is limited in the following cases:

- personality questionnaires,
- work sample tests,
- simulations.

Consideration must be given to a proper selection and introduction of trained controllers for the purpose of setting cut-off scores. It must be ensured that they are highly motivated to achieve the best results in tests. However, the

general fact that controllers are not in the same selection situation as applicants must be borne in mind.

One should also avoid a situation where ATCOs or trainees are constantly over-tested, they are not guinea pigs. Such a situation might also limit the access for testing them when necessary.

10.2.4 Policy Recommendations

The basis for defining cut-off scores should be clearly defined and as rational as possible, and should take the following into consideration:

- the validity and job relatedness of the test; only tests with a proven and sufficient high validity should be used for selection and hence a cut-off score can be applied;
- whether the test can only be used with reference to a norm sample or with reference to a criterion; if information on actual job or training performance is available, the cut-off score should be based on this criterion information (see example above);
- preliminary cut-off scores should be regularly checked and updated as soon as information for correction is available.

Regular Checks: Concerning the last bullet the following recommendations are made: It may be necessary to check cut-off scores from time to time and to adjust them if necessary. A check will be necessary if

- recruiting source has changed (e.g. higher percentage of applicants with third level education);
- the test has been changed (e.g. instruction, items, scoring procedures);
- the test battery has been changed (e.g. new test(s) to be included in the battery);
- training requirements have been changed (e.g. shorter training or higher performance requirements in certain aspects);
- insufficient performance of trainees (e.g. high failure rates, high number of trainees with marginal performance or poor training efficiency);
- the controller job has been changed (e.g. due to new technologies).

10.3 Separation of the Interview From Other Data Sources

There is an advantage in keeping the test results and the interview results independent. Keeping the predictors independent permits the application of a larger range of statistical procedures. For further explanation of these statistical procedures see Sub-chapter 7.1 of the TS.

10.3.1 Objective

Knowing the results of applicants in previous selection tools may unfairly influence interviewers. Training the interviewers may reduce potential bias but will not ensure its disappearance. Interviewers may be more susceptible for more bias. What happens when the interviewers are aware of the previous test results is unmeasurable.

Sub-chapter 8.1.7 already referred to the necessity of separating the interview from other data sources in order to establish an unbiased validity of the selection interview.

10.3.2 Policy Recommendation

This procedure results not only in impartiality of candidates and the establishment of the net unbiased contribution of each selection tool to the overall prediction of the criterion (e.g. success in training), but also in more cost-effectiveness in selection.

The net contribution of each selection tool is an important piece of information for the decision maker in selection management.

10.4 Weighting of Predictors

Other cost-effective improvements in selection procedures can be demonstrated by using **weighted** predictors of ATC success.

Crucial to the success of any selection system is the method of harmonising the parts of the system.

10.4.1 Objective

The objective of a multiple stage selection system is to acquire a variety of reliable data from people during successive stages of selection and **combine** these data so that persons with the highest probability of success can be chosen for training.

There are savings which can be derived in the selection system. These savings can be achieved by fully harmonising the different stages of selection. One desirable aim, from a cost-effective framework, would be to decrease selection throughput, while at the same time increasing the proportion of highly qualified candidates who undertake the training.

10.4.2 Practical Application

One of the ways of harmonising these selection stages is, for example, after an interview has been conducted, to derive a composite score from biographical data, scores from ability and achievement tests and the interview results, with which the final selection decision can be made.

Predicted values are of interest when the results are used to predict new data. One may wish to predict the mean score in training for all applicants with a given value of test score or to predict the value of success in training for a single applicant. Weighting the predictors provides a useful piece of information, which would not be available from other techniques.

Combining data from people during successive stages of selection in order to predict the highest probability of success in training is more cost-effective than using each predictor merely as a yes/no decision for the subsequent stage of selection.

Instead of increasing the intake of trainees, it could be worthwhile first to tackle the above-mentioned methodological deficiencies.

For further explanation see the TS, Sub-chapter 7.2 and EATCHIP (1998b).

10.4.3 Policy Recommendations

If we define an efficient selection as the ability of a system to achieve sustained high rates of success in fully qualified ATCOs, then, in order to keep with good psychometric practice, it is recommended to

- assign relative importance to each independent predictor variable, based on validation research;
- use the actual real scores of training or OJT for the validation and not the pass / fail;
- keep the test results and the interviewer(s) apart;
- combine the weighted independent variables (e.g. test and interview scores) into one success prediction, after the interview has been conducted.

As was shown in Sub-chapter 10.1 the method proposed also gives information on where to set cut-off scores in test distributions.

10.5 Sources of Error in Measurement

When referring to professional practices in test application and evaluation, one has to be aware of sources of error in measurement which may contribute to variations in a respondent's answer to a selection measure.

10.5.1 Temporary Factors Affecting Responses

Factors that are temporary but affect responses to a selection measure are, for example, a respondent's

- physical health,

- fatigue,
- level of motivation,
- emotional stress,
- general “test-wisdom” in how to answer or respond to a measure,
- readiness to complete a measure and
- physical conditions (heating, cooling, lighting, etc.) under which the measure is administered (Gatewood and Feild, 1987).

10.5.2 Temporary Factors Applying to a Specific Measure

Factors that apply to the measure as a whole are:

- respondent’s understanding of the instructions for a measure;
- respondent’s knowledge of specific tricks or techniques in responding to a measure;
- amount of practice available to a respondent for specific tasks required in responding to a measure (Gatewood and Feild, 1987).

Factors that apply to particular questions on a measure are:

- lapses of memory,
- fluctuations in respondent’s attention.

Other factors are:

- error of measurements (resulting from bad data, computational errors, transcriptional errors and typographical errors), range restriction and sampling error;
- “chance” element associated with specific items chosen for a measure;
- variations due to “guessing” or random answering, (see also Sub-chapter 10.6 below on “Guessing”).

10.5.3 Policy Recommendation

Longer tests usually provide freedom from one type of error of measurement and hence a higher reliability. However, the inclusion of irrelevant material in the content sample of the test items just for the sake of lengthening the test is not a justifiable procedure.

Regarding the effects of combining different parts of a test (e.g. just for increasing the reliability) see Sub-chapter 6.2 of the TS.

10.6 Guessing

The possibility of guessing the correct answer is an inherent phenomenon in a multiple-choice or true-false item test. When an item has a correct and an incorrect answer(s), a candidate may provide a correct response through guessing, without actually knowing the answer.

The possibility of guessing and the different strategies of behaviour in an attempt to guess the correct answer may significantly complicate the psychometric considerations. Guessing particularly needs to be taken into consideration when measuring a cognitive construct (Suen, 1990).

Test taker strategies, such as guessing, skipping all doubtful items, or skipping and then returning to doubtful items as time allows, can influence test scores positively or negatively, depending on scoring procedures, speed or the number of alternatives provided in multiple-choice item tests. Differential use of such strategies by candidates is unpredictable and may result in reduced test reliability and validity (CDSEPT, 1999).

Since the total score is usually the sum of individual item scores, the unknown probability of guessing could influence the reliability of the total score.

Guessing can be a serious form of distortion in a true-false item. If a test were balanced for true and false correct answers, shrewd candidates should be able to obtain a 50% score.

10.6.1 Practical Applications

Calculating the probability of guessing is usually not derived empirically, but is rather based on assumptions about how we think candidates (who do not have the necessary knowledge to select the correct answer) behave.

There is no evidence to support any formula of correction for guessing based on a random guessing model (Suen, 1990).

Test takers do not guess at random. Guessing may depend on the quality of distractors (the incorrect options in a multiple-choice item test). It may also depend on some unknown partial knowledge which may contribute to a higher or lower than chance probability of answering correctly. Based on partial knowledge, test takers may be able to eliminate some distractors. For example, in a conventional four options item, the random probability of 25% to answer correctly may be enhanced to 50% if thanks to partial knowledge, the candidate is able to eliminate two distractors (Crocker and Algina, 1986).

Lord (1974) suggested that the probability of a correct answer is less than random because distractors are constructed not only to be plausible but also to be attractive to those candidates who lack the necessary knowledge.

One can differentiate between variations due to “guessing” or random answering on the one hand and the “chance” element associated with specific items chosen for a measure on the other hand. The probability of guessing correctly is the likelihood that a test taker with absolutely no knowledge will still have a given probability of responding correctly. Because guessing behaviours are not sufficiently well understood and the value of the probability of guessing correctly is lower than pure random guessing, Hambleton and Swaminathan (1985) suggested referring to a ‘pseudo-chance level’ instead of guessing.

10.6.2 Policy Recommendations

Where candidates were instructed ‘not to leave blanks’ and guessing may have occurred, any assumption that there is equal chance for each option to be chosen may under or overestimate a candidate’s scores. Therefore, guessing-correction formulae may well be wrong for a given candidate and as a result are not highly recommended.

Ackerman and Kanfer (1993) described a formula of calculating in a multiple-choice spatial orientation test for *ab initio* trainee controllers with five distractors where the total score is equal to the number of correct responses minus .25 multiplied by the wrong responses.

From a practical point of view since (as was mentioned) the likelihood of random response due to partial knowledge is unpredictable, a generic formula for penalising in all tests would not be appropriate. Penalising procedures may actually require further (expensive) investigation into specific formulae for each item.

When penalising is impossible in the case of multiple-choice items with an unpunished guessing possibility, the values of item difficulty of about .50 or slightly higher are ideal for maximising the reliability (Crocker and Algina, 1986).

For further details on item difficulty see Sub-chapter 3.2 in the TS.

Guideline 40: Guessing

The possibility of guessing should be avoided as much as possible, or should result in a reduced score for false guessing. Penalising wrong answers may increase test validity.

When no other alternative exists and guessing is permitted, the information on the possible effectiveness of such strategies should be conveyed to all candidates in order to provide equal opportunity to present the best performance possible.

10.7 Rank or Critical Scores?

If a selection instrument measures a substantial and important part of the job validly and reliably, and provides adequate discrimination in the score ranges involved, persons may be ranked on the basis of their results.

However, if an instrument is constructed more in the manner of a training mastery test, in which the test taker is expected to answer all or nearly all of the items correctly, a critical score may be appropriate.

A critical score is also appropriate in situations where greater speed or accuracy cannot be reflected in performance because of equipment or process limitations. In that case, the limiting conditions should be considered in the design of the selection procedure (SIOP, 1987).

10.8 Re-testing

This STF has not found any conclusive (i.e. empirical) evidence regarding re-testing, re-evaluating, or reconsidering candidates for *ab initio* trainee controllers in the ECAC area or elsewhere.

Regarding re-testing, there are different findings and procedures depending on test systems, national differences, local unemployment rate and availability of the 'off the shelf' test. In some States the right of a test taker to take the test again is guaranteed by law. Therefore, individual States should make their own decision regarding the cost-benefit effectiveness of re-testing.

10.8.1 Practical Application

One way of calculating the cost effectiveness of re-testing is to calculate the cost per person for those who passed the test. This will be equal to the total cost of one test taker, multiplied by the total number of applicants who sat the test X times (e.g. twice), divided by the total number of applicants who actually passed training after taking the test X times (e.g. twice).

By comparing the results of 'cost per person passed' of different groups and of 'number of test written' one could possibly arrive at one reasonable conclusion regarding the policy of re-testing.

10.9 Review of Validation Procedures

It might be wise to have the validation procedures (with the technical manual) reviewed from time to time by one or more independent experts and their judgements recorded. This will not only serve as a quality control measure, which identifies deficiencies, but may also help in possible legal challenges later.

10.9.1 Practical Application

The review can start by looking at the predictive validity. If it is found to be too low, one could search for explanations by working backwards through reliability and item analysis.

This review may include the comparison of mean raw scores (in different elements of the test battery) of those who failed training against the mean raw scores of those who passed training. This may be followed by constructing expectancy tables based on actual raw scores (of the selection tests) of trainees who passed and failed. At the same time it would be noteworthy to check (with the help of item analysis) potential effects of learning between different parts of the test battery on the validity.

If extreme deficiencies were found, a review of the test development procedures and methods of item selection may be necessary.

Only high concordance (i.e. willingness and agreement to work together towards common goals) among selection, training and OJT sub-systems ensures a truly reliable and valid system of performance measurement.

10.9.2 Policy Recommendations

Re-examining reliability and validity of data may lead to modifications. Alternative models of prediction may be calculated (e.g. if the model contains six tests and some biographical data, there are different combinations of predictors which can be compared).

Ensuring the long-term integrity of the database is a daily task, not seasonal.

New selection instruments may have been developed. Existing commercial “off the shelf” selection tests may be of interest. When a test has been compromised and becomes freely available on the market to all applicants, it is rendered ineffective in predicting performance and appropriate new selection tests should be sought at the earliest possible opportunity.

Continuous inquiry into population changes that may affect future recruitment (e.g. video games and keyboard experience of candidates) is usually needed.

Guideline 41: Review of Validation Procedures

The review of a selection programme and its validation procedures should be conducted as an ongoing process.

10.10 Use of Test Catalogues

ATSP organisations would be well advised to choose commercial tests carefully. Commercial tests should not be selected out of test catalogues, usually offered by testing companies, on the basis of the label applied to the

test or unsupported claims in the test manual. Test labels can be quite misleading.

A full validation process in the local ATC environment is recommended before applying any available commercial test. Test names and unsupported claims in promotional test manuals are not evidence of construct validity (Tenopyr, 1987).

If a publisher packages tests that are to be used in **combination** for selection, the user should review the technical manual for the rationale for the specific combination of tests used and the justification of the interpretive relationships among the scores. For example, when achievement, general ability and interest measures are combined to suggest prediction of success, then supporting validity data for this combination should be given (CDSEPT, 1999).

10.11 Optimisation of Selection by Analysis of Age Brackets

Anecdotal evidence may indicate that beyond the age of 30, the age of the *ab initio* trainee controller is negatively related to success at Initial Training and at OJT.

10.11.1 Practical Application

A more professional practice will compare proportions (by age of persons in years) of those who applied to become ATCOs to the proportion of persons who subsequently graduated and were licensed for each age group.

The percentage of licences by age group (e.g. percentage of all licensed ATCOs who were 24 years old when they applied out of all licences) divided by the percentage of applications by year (e.g. the share of the 24 year olds out of all applicants) can give a deviation from benchmark. Should the value of the deviation be more than 0, then one can conclude that this year of age is contributing more than its fair proportion to the total of licences. In other words, the ‘return on investment’ (in interviewing, conducting security checks, training, etc.) is positive for this age group.

One could also calculate the value lost (in money) to the ATSP organisation by continuing to reject candidates above 25.

For a fairly narrow age range (e.g. *ab initio* trainee controllers are usually restricted to 18-25 years old), frequency distribution may be utilised for each month of age (e.g. 19 years 3 months) in analysing this important variable. While broadening the age range, it may be advisable not to present a straight line of prediction, but rather to fit a smooth curve to the points using graphical methods. This way of presentation enables the smoothing out of irregularities in the data. Refinement of scores may be applied, but any interpolation and extrapolation of a transformed chronological age should be carried out with great caution regarding the twelve-month period.

Some scores (e.g. in risk taking behaviour scale of a personality questionnaire or in some psychomotor ability tests) increase with age. 'Partial' correlation may be used in such a case to obtain a measure of relationship between two variables (e.g. risk taking and decisiveness) with the effect of age removed.

10.12 Miscellaneous

10.12.1 Computerisation of Test Battery

In many instances, the switch from paper-and-pencil to computer-based tests will require additional evidence that relevant standards (e.g. reliability) have been met in the new testing mode (CDSEPT, 1999).

10.12.2 Testing Locally in Different Regions

Those responsible for national testing programs should ensure that individuals in local regions who use the test scores are properly instructed in the appropriate methods of administrating and interpreting test scores (CDSEPT, 1999).

The principle of transparency of local or national selection procedures ensures not only a higher reliability of the selection tools in use but also a better standing in would-be legal challenges.

A provision should be made that the interview panel will include members from different regions of a given State in order to avoid a bias.

10.12.3 The Responsibility in Case of Outsourcing of Selection Service

As was mentioned in Chapter 4, "Validation Issues", with regard to test developers who should supply the required information on validation, the ultimate responsibility for appropriate test use lies with the user who requires the test results for decision-making purposes. The same applies when the whole selection process is contracted out. This responsibility cannot be abrogated by purchasing services from an outside provider.

For the appropriate interpretation of reliability data and norms supplied by a contractor or some professional provider see also Chapter 6 of the TS.

Page intentionally left blank

11. DECISION-MAKING PROCEDURES

This Chapter outlines the underlying notions and gives recommendations with regard to decision-making in selection.

11.1 Guiding Principles of Decision-making

It has been stressed in the preceding Chapters that one of the main goals in testing or assessing applicants in the selection process is to furnish precise estimates of the applicants performance, behaviour tendency or personality which are relevant (valid) for performance in controller training and job.

At several stages of the selection process, in particular in a multi-stage selection, decisions have to be made as to whether an applicant should be considered for further testing or - in the final stage - should be accepted as a candidate for intake in an *ab initio* training course.

The importance of the outcome of such decisions for both, the organisation and the individual is obvious. It is therefore clear that not only the measurement procedures and the predictive validity need to be evaluated but also the outcomes of personnel decisions. In fact, from this viewpoint, all measures taken by means of tests or assessments made (e.g. in an interview) are purely technical components of a system which is designed to make decisions.

11.1.1 Discriminative Efficiency of the Test Battery

Good predictive validity of the procedures used is a necessary but not sufficient condition for arriving at decisions.

If a decision maker knew exactly how well each applicant is qualified, and what therefore the outcomes would be for each applicant in controller training and job, no decision problem would exist. However, this exact information is not available.

From the practical view of decision-making, the number of correct decisions made by a decision board on the basis of a test data or assessments is essential. Hence, the discriminative efficiency of a test or test battery, i.e. the extent to which a test / test battery is able to accurately separate the different outcomes (between ‘valid positives’ and valid negative’) is of more importance to decision makers than the correlation between test scores and performance criteria (e.g. training performance).

The better a test battery discriminates between ‘valid positives’ (candidates that would be successful in training) and valid negatives (those who would not be successful) the higher is the probability that a decision to accept / reject an applicant results in high number of actual successes in training (‘**valid**

positives') and small numbers of failures ('**false** positives'). For details see also Annex F in EATMP (2001d).

Guideline 42: Discriminative Efficiency of the Test Battery

For decision-making, information on the discriminative efficiency of tests and assessments is most important and necessary to make predictions on the outcomes of the decisions to be taken.

The extent of accurate separation between valid positive and valid negative applicants by the use of the test battery is essential information for a better-than-chance decision.

11.1.2 Role of the Decision Maker

The willingness to assume responsibility for making selection decisions whilst not having perfect information available is a necessary, but certainly not a sufficient qualification of decision makers. Decision makers need to be willing also to utilise any valid information that will enable them to forecast the outcomes of their decision more accurately and make better-than-chance assignments or decisions.

Data from application form, tests and other procedures can provide this information.

The role of the decision maker is then

- to integrate the information;
- to collate the weighted predictions with regard to the possible outcomes of the decision;
- to arrive at a decision.

Decision makers need therefore a clear strategy or rules for arriving at decisions which indicate the steps to be taken in the light of the information they have available (Wiggins, 1973). The role of decision makers is then to consistently apply this strategy and follow the prescribed rules.

For documentation of decision rules see the selection guidelines, Sub-chapter 5.2.9 (EATMP, 2001b).

A simple personnel decision strategy might be (in the case of just one test as the database): "If the applicant's score in the test is above a certain test score "x" (e.g. a cut-off), accept the applicant as a candidate for training. If the test score is below test score "x" reject the applicant."

The situation is more complex if more than one piece of data or information is available and needs to be integrated. This more common situation will be tackled later.

Guideline 43: Role of the Decision Maker

Decision makers need to be willing and able to arrive at their decisions.

They need to utilise relevant (valid) data and other information. This data should be made available to decision makers and a clear strategy should be developed which gives guidance to decision makers.

The decision strategy should be applied and the rules should be followed consistently.

11.1.3 Values of Decisions

Sufficient emphasis should be given to the fact that the possible outcomes of any decision are not of equal importance or value over time.

The value of a decision to accept or reject an applicant might be disturbed due to some of the following value factors:

- safety of the service,
- filling open vacancies (e.g. manpower needs),
- reaching selection quota,
- homogeneity in groups for intakes (e.g. student course),
- length of waiting period for new candidates until course starts,
- affirmative action policy,
- public relations issues.

The list is not exhaustive, but it is obvious that these value judgements can impact on the decisions taken especially in ‘borderline’ or somehow ‘doubtful’ cases. Due to external influences, decision makers will not always be able to maximise the ‘gains’ and minimise the ‘losses’ for the organisation they represent.

Guideline 44: Values of Decisions

A policy as part of the decision strategy should be adopted which clearly states the guiding principles of selection decisions, their values and how the decision board should proceed in certain cases.

11.2 Types of Decision

The following three types of decision are possible:

- **(Single) Selection**: Only one position type in the job and/or training scheme is available and decision makers are allowed to reject applicants.

  This is the situation which exists in most States with regard to *ab initio* trainee controller selection. Only one type of job (controller) and/or training is available and the decision is based on a single test battery relevant for this type of job.

- **Multiple Selection**: More than one job position and/or training scheme is available and rejection of applicants is allowed. Candidates are either assigned, based on their test results, to an appropriate job / training schemes or rejected.

  This situation does not exist, at the present time, although multiple selection would in principle be possible in ATC if (a) a test battery would allow for multiple assignment (e.g. a multiple test battery) to different controller jobs and (b) the practice in MP would allow an early assignment of candidates to different jobs on offer and (c) different training schemes could be offered.

- **Classification**: More than one job position / training scheme is available and all applicants are retained. All candidates are assigned to a certain job or training (e.g. ATCO in a low density tower or other job such as Aeronautical Information Services [AIS]).

  This situation rarely exists. This strategy has similarities with the placement system in Poland, where other jobs within the Polish Air Traffic Agency may be considered or in France where applicants for a job in ATC are trained and during the course are placed according to their training progress / performance at a certain job position. Nearly all trainees are retained (see also Sub-chapter 5.3.5).

In the following only the first selection strategy will be further considered.

11.3 Strategies in Decision-making

11.3.1 (Multiple) Regression Model

Tests and other assessments are sought which correlate highly with a certain performance criterion in training or in the job and show low intercorrelations. Using the statistical method of Multiple Regression (see also Chapter 5 in the TS) a combination of these measures yields a single composite score for each applicant.

The different measures (e.g. test scores) are weighted in terms of their correlation with the criterion and their independence from all other measures (other tests or interviews). The composite score is the sum of the weighted test or assessment scores.

Guideline 45: Regression Composite Scores

If the relationship between predictors and criterion measures used in selection decisions is described by regression equations, the associated standard errors of estimate should be applied to establish the cut-off in the composite score.

Prediction Optimisation: An optimised selection strategy involves the optimal weighting of a set of independent predictors to maximise the correlation between predicted and actual performance (that is, the criterion).

Example: Tests which have low correlations with the criterion and/or are highly correlated with other tests in the test battery are given lower weights (and vice versa).

This straightforward procedure addresses the following issues:

- a test battery, which consists of tests, which are not or are insignificantly correlated with the performance criterion will almost certainly have no weight in the regression equation and no composite score can be calculated;

- a test battery which consists of a number of redundant tests, i.e. tests which measure the same ability will almost certainly gain small weights or might even have no additional weight in the equation (are superfluous).

This indicates again the importance of having validated tests available in selection, since otherwise the decision cannot be based on additional information, which helps to make better-than-chance assignments.

In addition to this, the second bullet point highlights the fact, that a highly redundant test battery will not give more information to decision makers than a test battery which consists only of those tests which independently contribute to the prediction of performance. From this point of view, the former test battery would yield unnecessary information, which is useless for decision-making.

The multiple regression model has a side effect which might be undesired. It is a compensatory model, that is, although the statistical weights are constant for all test takers, the test scores (which are weighted) will vary widely from test taker to test taker. Hence, it is possible that test takers with different profiles of test scores obtain the same or nearly the same composite score. This is due to the fact that a low test score in one test can be compensated in this model by higher scores in another test.

Example: It is assumed that the test battery for *ab initio* trainee controllers consists of two ability tests, a test for 'perceptual / psychomotor abilities' (test X) and a test for 'spatial comprehension' (test Y) both having different but positive correlation with a performance criterion in training. The multiple regression of both tests on the criterion would yield a weight of .40 for test X and .10 for test Y.

Two applicants with different test scores in both tests (applicant A: test X = 5 and test Y = 20; applicant B: test X = 10 and test Y = 0) would then obtain the same composite score of 4.00 each. However, both applicants are clearly not equivalent in their potential to be controllers. In fact, applicant A would be more suitable than applicant B.

A model which would allow to set minimal levels of test performance would in this case be more appropriate (see also Sub-chapter 7.2 in the TS).

11.3.2 Multiple Cut-off Model

The alternative to this model would be the "Multiple Cut-off Model" described next where **minimal** levels of test performance can be set and all candidates below this level in one or more tests / assessments would be rejected.

This approach is fully in line with the method described in Sub-chapter 10.1 concerning cut-off scores and the considerations that should be given to different methods in doing this.

The multiple cut-off approach is the more conservative approach of the two discussed here. The determination of optimum cut-off scores as was described in Sub-chapter 10.1 becomes laborious in the case of test batteries with more than three tests.

The solution (which is often used in setting multiple cut-off scores) is to simply set the cut-off at the mean of the test scores from a norm sample. This practice cannot be recommended since the cut-off is not set following empirical checks and no consideration is given to external influences such as quality of applicants and the predictive errors that will be tolerated (e.g. failures in training).

The situation would be even more inefficient if the tests used are highly interrelated (redundant measures) and/or are not valid predictors of a performance criterion but all tests are used to screen out applicants based on a cut-off score which was set intuitively. It must be expected that

- a too small percentage of applicants will be selected;
- a high number of applicants are rejected which would be able to pass training ('false negatives').

A combination of both models, the multiple regression and multiple cut-off model would be the best solution. After applicants have been screened due to cut-off scores in important and valid tests (e.g. tests where applicants should

have reached a minimum level of performance), then for the remaining applicants a composite score could be calculated indicating their test performance across all tests / assessments. This 'multi-stage strategy' is described below in more detail.

11.3.3 Selection Decisions in Multiple Stage Selection Systems

A multiple stage selection system is characterised by a sequence of tests and interviews where decisions are made after each sequence whether or not applicants should enter the subsequent stage of the process. In the following example the preceding stages in pre-selection for admittance of applicants to more formalised selection testing are left out.

For example: The first stage in the process could be a paper-and-pencil or computer-based test battery on general and specific abilities. A decision is made on the basis of test results achieved following a multiple cut-off strategy. The second stage could be psychomotor / perceptual tests or other, more complex tests and a decision is made whether applicants should enter a final stage, an interview.

Once the test scores have been used for this purpose, they can still be considered useful and can be used in combination with the interview results and biographical data to derive a composite score (from regression) upon which the final selection decision is made.

The final decision in this multiple stage strategy based on the composite score can be taken in one of the following ways:

- either another cut-off score is applied to the composite score and all candidates are accepted to fall above this score or
- candidates are selected from the top (which would be the more conservative approach).

Guideline 46: Decision Strategies

A decision strategy should be designed which ensures that decision makers can rely in their decision on an optimised combination of predictors. A composite score should be derived from all relevant information and from all stages of a multiple stage selection system for choosing the most apt for controller training and job.

It might be necessary to counter the side effects of a regression approach by setting (multiple) cut-off scores in important predictors where applicants should achieve minimum levels of performance.

11.3.4 Derivation of Composite Scores

Combining results in different stages of a multiple stage selection system is only justified, if tests or other methods can be combined in a way, that in each stage applicants who meet the requirements in regard to

- later selection stages,
- success criteria (training success and/or job performance) or
- both

are selected.

The first bullet refers to a screening procedure, which ensures that applicants who have a higher chance for success in later selection stage(s) are pre-selected. The second bullet refers to a screening procedure, which is normally applied to select candidates who have a higher chance in regard to success criteria.

11.4 Use of Expectancy Tables

Expectancy tables can give additional information to decision makers on the probabilities of outcomes (e.g. success in training) which is associated with certain categories of test scores or composite score derived from a regression equation (see also Sub-chapter 10.2.3).

12. RECRUITMENT, SELECTION AND TRAINING DATABASE

This Chapter outlines the principles and requirements of a standard database for *ab initio* trainee controller applicants, trainees and licensed ATCOs.

12.1 Rationale and Purpose of Database

The recruitment, selection and training system is charged with the responsibility of finding, choosing and training *ab initio* trainee controllers. The objective of MP is to provide a sufficient number of appropriately qualified people, on a timely basis, for the work as ATCOs.

In order to understand, control and improve the performance in recruitment, selection and training data on the whole process is required. A database can be a powerful tool upon which certain actions can be initiated to improve the performance of the recruiting, selection and training activity.

Example: In order to understand and improve the success rate in training, applicant data from preceding recruitment and selection stages and from training must be available. The effectiveness of a current recruiting, selection and training is monitored, validated and controlled by analysing the data gathered through the database.

The general purposes of an integrated database for a recruiting, selection and training system are

- to gather important data,
- to analyse data,

in order to

- report on achievements and past records,
- maintain the selection tools and methods,
- validate the system,
- predict future performance.

12.2 Objectives and Requirements

The overall objective is to establish a single, reliable and complete database, which allows all functions to be completed. All requirements outlined in the following sub-chapters must therefore be fulfilled no matter who maintains the database.

The functions outlined may be contracted out to a service provider, but the database must include the information necessary to conduct all activities.

12.2.1 Performance Tracking

Performance tracking for purposes of control and reporting of all possible achievements is a basic function in any recruitment, selection and training database system.

Objective

The objective of performance tracking is to establish and maintain a single, reliable personnel performance tracking system to capture and record the performance and disposition of each candidate / trainee at each stage of recruitment, selection and training.

A simple example of performance tracking at the selection stage is shown in Table 2. For an example of performance tracking at the training stage see Table 3.

Table 2: Example of performance tracking at the selection stage

Selection Period: 25/01/93-23/07/93

Selection Stage	**Took the test**	**Failed**	**%**	**Passed**	**%**
1) Paper Sift	65	26	40%	39	60%
2) English language Test	39	16	41%	23	59%
3) Psychometric Test	23	8	35%	15	65%
4) Interview	15	4	27%	11	73%

Overall pass rate: 11 out of 65 = 17%

Table 3: Example of performance tracking in training

Training Period: 01/02/93-31/08/98

Course N°	**Start Date**	**End Date**	**Intake**	**Failed**	**Recoursed**	**Resigned**	**Completed**	**Success Rate**
17	1.2.93	31.1.95	15	3	2	1	9	60%
18	1.4.94	12.3.96	15	6	1	1	7	47%
19	1.9.96	31.8.98	15	Pending				

Requirements

The requirements for the database to track performance depend on the level of detail required in a report.

The applicability of standard software which provides the necessary descriptive statistics should be possible in this case.

12.2.2 Maintenance of Tests and Assessment Procedures

Test and assessment maintenance is a necessary basic function to assure the quality, reliability and validity of applied methods of selection.

Tests, exams and other assessment procedures (e.g. behavioural observation scales, performance ratings) applied in Initial Training and OJT need to be maintained in order to assure the quality in application and to check the reliability and validity of these procedures (if applicable).

This function is currently widely neglected as outlined before. However, the improvements that are possible and necessary in the whole process of recruitment, selection and training would not be complete and/or might fail to achieve its objectives if this important area is left out.

Objective

The objective of maintenance of tests and other assessment procedures is to control and assure the quality of applied selection procedures and tests, to check regularly their reliability and to validate them, if applicable.

The same objective can be applied to certain assessment methods and tools in training.

Requirements

In order to carry out test and assessment maintenance, a database containing test data from candidates during successive stages of selection, training and on the job (e.g. measures of job performance) is required.

The amount and type of data required will vary depending on the type of maintenance procedure to be applied. For example, for validation purposes detailed test data, data on performance criteria in training and on the job is required.

Further technical details on these issues are given in Chapters 3, 4 and 6 of the TS.

12.2.3 Development of Tests and Assessment Procedures

Test development is an advanced function of a database. It may concern tests applied in selection or tests and other assessment procedures applied in Initial Training or OJT.

Objective

The objective of test development is to initiate item analysis for in-depth test control and maintenance and to develop improved test versions and/or parallel versions of tests.

A similar objective may be defined in regard to the development of tests and assessments in training.

Requirements

Detailed data on test items needs to be collected for each candidate or trainee to meet this objective.

The database may also contain item information on questions used in written tests at the end of each phase in Initial Training. This information may be used to generate tests which show similar characteristics due to item difficulty and content (see also Chapter 3 in the TS).

12.2.4 Integration in HRM Database

MP is an important HRM function. Systematic attempts to forecast manpower demands rely *inter alia* on the availability of relevant data from recruitment, selection and training.

Objective

The objective of integrating the recruiting, selection and training database in a more general HRM database is to ensure that data relevant for the MP is available and can be used.

Requirements

The data requirements will depend on related policies to MP and management.

The database must be easily and reliably uploaded to standard packages in use (e.g. statistical packages, graphical packages), see also EATMP (2001e).

12.3 Database Structure

Table 4 displays a list of items in a standard ATC personnel database for basic requirements. ‘A’ between brackets stands for advanced requirement.

Depending on the legal framework to be applied, detailed results of each test, data from biographical questionnaires, personality inventories and academic grades achieved during training should be stored. Otherwise, they should be destroyed after use, or after a certain time period.

Table 4: List of items in a standard ATC personnel database

APPLICATION AND SELECTION DATA	TRAINING DATA	COMPETENCY CHECKS AND REFRESHER TRAINING DATA
Candidate identifier	**Initial Training**	Last medical check date (A)
Screening status (e.g. paper sift mark)	Training Course number	Proficiency check date(s) (A)
Family name	Initial Training qualification date	Proficiency check score(s) (A)
Given name	Training score(s) [basic radar control, navigation, IFR non-radar, meteorology scores, etc.] (A)	**Refresher Training**
Application date	Resist in exams (A)	Normal rating refresher
Application city (if applicable)	Recoursed (A)	Emergency / unusual incident refresher rating
Gender	Circumstances of recoursing (e.g. sickness, failure) (A)	
Date of birth	Number of additional hours on simulator (A)	
Nationality	**On-site Training**	
Languages	Training scores	
Address	Number of hours at on-site simulator and OJT (A)	
Related work experience	Student assigned to number of trainers (A)	
Education (number of years completed in secondary school, university)	Student assigned to another shift / watch (A)	
Educational documents in relevant subjects	Student assigned to another career (A)	
	Qualification / Rating	
Selection Data	Aerodrome control rating	
Ability test date	Aerodrome control rating date	

Table 4 (cont.): List of items in a standard ATC personnel database

Ability test version number	**Air movements position**	
Raw score of ability test	Surface movements position	
Answers for question 1...n in the ability test (A)	**Aerodrome radar position** (not yet officially applied)	
Sub-parts scores of ability test (A)	**Approach control rating**	
Standard score of ability test (e.g. percentage, STANINE, t-score) (A)	Approach control rating date	
English language test results	Position date	
Interview date	**Approach radar control rating**	
Interview version number	Approach radar control rating date	
Percentage score of interview	**Approach precision control rating**	
Sub-parts scores (motivation, stress, communication, etc.) (A)	Approach precision control rating date	
Weighted score from test and interview (A)	Position date	
	Area control rating	
Medical / Security	Area control rating date	
Medical conditions (e.g. eyesight)	Sector / Position date	
Medical status (‘pass with limitation’, ‘pending’, etc.)	**Area radar control rating**	
Security status	Area radar control rating date	
	Sector / Position date	
	Number of training hours on position / sector	

12.4 Practical Application

Analyses of data in a complex database have extensive practical applications.

12.4.1 Validation of Recruiting, Selection and Training Practices

It would be crucial to the further investigation of system efficiencies, that the standard database contains the criterion data (e.g. measures of performance at OJT, e.g. “Aerodrome Control Rating” and/or job performance, e.g. “Proficiency Check Score”). Without this data it will be difficult, if not impossible, to validate the recruitment and selection system.

For example, can biographical and educational factors (e.g. age, grades earned in certain subjects, such as mathematics) which predict success in training be used to target recruitment efforts? This can in turn generate cost benefits and improvements in the whole selection process (see also EATMP, 2001d).

Of more than a little concern is the paucity of performance data in the database about persons who had either passed or failed the training. If the database contains only “pass / fail” disposition for candidates at training, or OJT, the application of appropriate validation procedures is severely limited. The direct result of such data deficiency is to reduce the power of the system to enable predictions.

In order to validate a selection system, performance in training and/or in the job needs to be quantitatively linked to performance in selection tests. Wherever possible, performance should be measured by numeric measures and not by pass / fail dichotomy (see also Chapter 5 in the TS).

12.4.2 Manpower Planning Impacts

The database can stimulate management to look more closely at training administration practices and to modify them.

For example, the data may be used to target potential training problems (e.g. variability in the amount of training time required to complete training and pass rate in relation to the specific OJT location) and can lead to improvements in training administration practices.

Of considerable concern to the MP authority is the fact that the system must provide, as early as possible, data to determine how many persons will eventually be licensed if a certain number of trainees enters the training course. It will also be useful to be able to update forecasts at each stage in training, for further discussion on MP forecasting and modelling, please see EATMP (2001e).

The database can be used to assess relationships between measures of OJT and Initial Training grades. All working positions necessary to become a fully

qualified controller and which are subject to a licence should be mentioned in the database.

The identification of bottlenecks that prevents the system from producing the required number of licences within the required timescale, as well as the identification of possible backlogs in training is of very great importance.

12.4.3 Database Application

Data will usually be transformed into information by posing questions relevant to certain objectives of the manpower system. These questions are then interpreted in such a way that appropriate statistical analysis of the data will supply answers.

A method to obtain statistical information using a well-established statistical package should be provided.

12.5 Database Management

12.5.1 Database Maintenance

Databases need maintenance, preferably by **one** person, provided it is legally possible. Often it is found that a database, for a variety of reasons, has fallen into disrepair.

Examples: Upon commencing the work of analysing, 'suddenly' files have to be reconfigured and some data inputs cross-checked and re-entered. Other files 'surprisingly' need to be reformatted, because unknown persons have opened and modified certain data inadequately. Anomalies, such as corrupted files are found. Correct versions of records need to be regenerated. Corrections that had been updated need to be replicated. Raw data entries need to be reviewed and recreated.

If the database is to become the powerful HRM tool that it can be, then care of it by a dedicated authority is mandatory.

12.5.2 Database Integrity

There should be **one** database, or one master file covering the whole period from candidate application to licensing as a professional ATCO. It is not recommended to store data from different stages separately, provided it is legally possible. It will be unfortunate if the database was unable to attach, for example, interview scores entered on one file to the file holding the performance at training.

Consideration should be given to the following factors:

- **The integrity of the database:** The integrity of the methods and procedures used to input data from data records such as application form,

test scores, student performance at training, etc., must be checked on a regular basis.

- **Sampling:** As a further assurance of the reliability of the database, the person in charge of the database should acquire from time to time a selected random sample of the 'actual' numbers. For example, scores from a training course shall be tracked back to their raw data files regularly to check the integrity and completeness of the data.

- **The long-term integrity of the database:** It is important to assure the long-term integrity of the database by assuring completeness of the data on a case-by-case basis.

Different alternatives for database formats exist. Depending on compatibility with current technical environment, the user should decide, for example, between files that can be read by the word processing or database software and/or other operating systems in use.

Page intentionally left blank

13. FUTURE TRENDS AND STEPS

13.1 Impact of Future Technological Advances on Test Development

This Chapter does not give Policy Recommendations that should be followed, but provides observations on events that are likely to happen in the future.

However, projecting current trends into the future is an ancient fallacy and nobody has or ever will have all 'the' correct answers to the questions that have been or can be addressed.

This Chapter does not aim to describe future ATC in terms of science fiction, but rather realistic reflections on the foreseeable job future and on a few possible impacts on future test developments. However, it still remains to be seen whether the technological developments described in this Chapter will lend themselves to solutions in selection and test development.

13.1.1 Pre-selection of Candidates

New technologies will have a big impact on the way *ab initio* trainee controllers are selected and trained.

A major issue in selection is, for example, pre-screening from the mass of raw applications hitting selection offices. Normally, administrators can sort this out by pre-selecting candidates according to basic requirements (e.g. age, language or education requirements), which is to some extent a straightforward procedure but involves also a detective sense and 'a nose for frauds'.

The question here is whether these decision processes can be programmed (e.g. on the Web). They are not programmed yet but may be programmable one day with the impact that future selectors might be dispensable.

13.1.2 Tests in Final Selection

The idea of customising tests, or using personalised tests, pre-dates the computer. But with paper-and-pencil versions of tests this opportunity was severely hampered.

Emerging trends in test development are about to change that. Current test developments clearly show some common features and trends:

- Tests will in most instances be computer-based.
- Computer-based tests allow for dynamic / interactive and adaptive testing. Some of the recent developments, for example, allow for performance monitoring during testing, that is, they are able to analyse test taker's performance during testing, adjust the type and/or the difficulty of items

presented and thus establish the upper and lower limits of the test taker's performance ability (see also EATMP, 2000a).

- Computer-based tests allow to measure and analyse response times on item and test level as well as error patterns and hence open new possibilities for multiple measurement / ability testing.

- Future tests may even include a verification routine in their software which, for example, stops the program after a certain date and replaces the test with a parallel version, or automatically changes the test version once the system realises that the person has already taken this test in an earlier application, provided it is detectable.

We can also expect the introduction of more tests of the type that simulate the actual work environment as for example:

- tests where candidates perform a variety of ATC-related tasks (e.g. communication) while being subjected to unexpected interruptions;

- tests measuring motor skills and not only cognitive skills;

- tests measuring important capabilities by using auditive media.

It can also be expected that the efficiency of pre-training screening programs is reviewed in the light of new findings based on new test developments in the direction mentioned.

The use of computers as a convenient medium for presenting test items asks for specific measures to avoid that applicants highly skilled in the use of computers ('computer kids') do not have advantages in computerised testing (e.g. by using simple to handle input devices, or allowing for enough practice in the use of input devices).

13.1.3 Three-dimensional Displays

Some of the interesting future advance for test development in the ATC domain may occur in relation to possibilities of displaying 3-D tests. Some possibilities are mentioned below.

- **Holography** - Replicas of the real world, on display and in computer memories that look 3-D are actually flat and rely on tricks to fool the eye by creating 3-D moving images. The current stage of holography development may provide a stable light-sensitive material needed to see an image from every angle in improved holograms (Berg et al., 1996).

- **Virtual Reality** (VR) or what is now known as "Immersive Virtual Reality" (IVR) promised a freedom of movement in a virtual-reality world. Development in VR genuinely produced a 3-D image, which observers can see from any position without any additional equipment (Downing et al., 1996).

However, the fidelity of animated computer graphics depends on the speed and capacity of computers to calculate and display the geometric rudiments of 3-D objects (polygons) to be displayed. This is critical in particular for Web-based testing. Once the technical problems of the 3-D display are solved, these devices might in the future be used in displaying tests for *ab initio* trainee controllers, which will be much closer to the reality (and ability) of the work of a controller.

Future technological developments might enable designers to build and test a prototype ATC environment without updating strips or communicating with aircraft. The technology could also be useful for creating environments in which to learn skills such as ATC that would be expensive, dangerous or even impossible in real life.

Latest versions of personal computers can create and display 3-D images which can be put on the Internet. This might create the demand for a home virtual reality system.

13.1.4 Using High Quality Sounds

Patents have been filed in the US for a new way to produce sound by combining two ultrasonic sounds (sounds higher in pitch than those that humans can hear) into a single, audible one (audible sound) with a far higher quality (The Economist, 1996).

There is still a lot of prototype building to do, but this development could capture the attention of developers of audio tests for ATCOs, which is one of the domains in testing where further developments would be gainful.

13.1.5 Compact Disk ROM

The introduction of Compact Disc Read-Only Memory (CD-ROM) offers new possibilities in test delivery and application (and marketing of the ATC job) which could be beneficial.

Test scenarios and instructions can be presented more easily and in an interactive manner with integrated video clips supported by sound or speech.

Future CD-ROM will be even more interactive than today, with a scope to interrogate images or text.

13.1.6 Internet Technology

Information technology, and the Internet in particular, could make information on selection procedures and occupation opportunities abundant and could help to serve it up faster (see also EATMP, 2001d).

However, it also makes selection tests more transparent, allowing candidates to prepare themselves more easily. The basic assumption of fairness in testing will be called into question by these opportunities.

Although information technology does not require the basic rules of psychometry to be rewritten, this does mean that psychometrists must be more on their guard and develop counter measures. The following difficulties have to be taken into consideration:

- The sheer volume of material on the Internet makes it difficult to patrol cyberspace for unwanted publication of selection tests.

- A future code of conduct for Internet providers would probably not be enough and/or would be too expensive for small Internet providers to go through all the material and to check what is carried on the Net.

- A clear policy on the use of the Internet and on the material distributed by it is currently not in sight as an issue to be solved on an international basis.

- The Internet has become a means to avoid or even break national and international laws and/or laws on intellectual property. Because the Internet whisks material around the world, it is harder to police than a national television channel or a newspaper.

- Hackers challenge the security of Web-based testing and users will need to apply more advanced methods to defend their tests.

Internet service providers may be forced to take responsibility for what they knowingly carry on their sites. However, the Internet will probably always allow the really determined user access to illegal material, such as, for example, copyrighted tests for *ab initio* trainee controllers. The FEAST project tackles those challenges and their possible solutions.

13.2 Future Trends in Increasing the Effectiveness of Training

The following trends in controller training can be foreseen to increase success in training and hence influence selection tools:

- adoption of a “training-for-success” philosophy, similar to what has been done in Switzerland, including (a) improvements in the quality and selection of OJTIs, (b) adopting computer-aided learning, (c) using simulators to supplement OJT and (d) individualising instruction;

- enhanced assessment and testing methods and procedures.

There is a certain area which needs to be looked at in particular in recruitment and selection:

- Recruitment, selection and in particular training is in the majority of the ECAC States geared towards students who come directly from secondary level education and in most cases without any other professional experience.

However, due to the conditions in the job market of the 2000s, a much greater proportion of students enters controller training with higher education levels or professional training and experience. The mixture of educational levels and experience is one of the challenges training institutes have to face and have to adapt to.

13.2.1 Enhanced Simulation

Enhanced computer simulators (and CBT packages) will permit trainees to be trained anywhere in the world, as simulators can already be programmed to replicate the characteristics of any operational sector. Remote training and e-learning are already a fact as far as Initial Training is concerned.

However, in order to maximise the use of simulation, particularly for the major centres, there will be a growing need for site-specific, high-fidelity training replicating the operational environment more exactly.

In future, trainees will face actual simulation at a much earlier stage in training.

The greater use of simulators will enable training of controllers anywhere in the world.

13.3 Future Trends on the ATC Job

13.3.1 Traffic Increase

According to the baseline scenario of the Specialist Panel on Air Traffic Statistics and Forecast (STATFOR) there will be an annual rise of number of flights in the EUROCONTROL Central Flow Management Unit (CFMU) area between 3.7% and 4.0% between 2002 and 2005. (Vandenberghe et al., 2001)[3].

A forecasted increase in air transportation will not only be translated into added labour requirements in ATM but will lead to increase in the political and environmental pressure for reforms in ATM.

13.3.2 Technological Evolutions

The latest ATC hardware and software has not yet altered the core of the controller's job which is to make priority decisions so that aircraft are sequenced and properly separated along "tunnels in the sky", including airport approach or departure paths (Hardeman, 1994).

While the new technology can aid decision-making, it is not yet (and probably will never be) a substitute for sound judgement or reassuring human voice, especially when handling emergency situations.

[3] This forecast was prepared before the tragic events of 11 September 2001

However, technological inventions will result in more effective display and processing of information, in simplifying procedures, such as aircraft hand-over and replacing voice communication between pilots and controllers with electronic communication (Stix, 1994).

Controller workload associated with the tasks a controller has to perform to handle flights through a sector will be a key issue and a major limiting factor and will determine sector capacity (number of aircraft that can be handled per time unit).

Wherever possible, controller tasks will be 'eliminated' or performed more efficiently, so that the controller can concentrate on core ATC and thus deliver increased system capacity (Brooker, 1996).

The work in the following technical evolutions currently under way within EATMP will have a major impact on the controller job, its requirements and the management of the system:

- Automatic Dependent Surveillance (ADS),
- Airborne Collision Avoidance Systems (ACAS),
- Satellite Navigation and Global Positioning System (GPS),
- Communication, Navigation and Surveillance (CNS) / ATM.

For further discussion of the effects of future ATM systems and their impact on the controller job, please see Fleming (2001).

The next generation of operational ATC systems will use more intelligent Human-Machine Interface (HMI) systems (EATCHIP, 1998d). The amount of information to be presented to the controller team in the working position will considerably increase the workload, especially in high density situations. In a 'stripless system' environment, sector traffic lists, message windows, dynamic formatting of the radar identification labels and radar information could be presented on one single medium (Garcia-Avello and Swierstra, 1997).

It is probable that more intelligent tools will present a higher information content, and subsequently it will be more difficult for controllers to identify malfunctions (Garcia-Avello and Swierstra, 1997).

This will require the development of adequate training programmes to provide appropriate skills for working with the automated tools. It will also be necessary to adjust controller selection tools.

13.3.3 Pace of Change

The ATC world is, as other high technology environments, marked by rapid changes that are difficult to cope with.

The readjustments necessary between one technological period and another are challenging and sometimes painful for organisations and people. The increase in technical efficiency has taken place almost faster than can be dealt with, for example, in controller training and in selection.

The pace of change that can be kept up with by developing countries maintaining increasingly sophisticated systems and equipment is one of the major concerns in future ATC (Proulx, 1996).

13.4 Future Trends on HR Planning and Management

ATC has experienced several technological changes since its early days and each new generation of technology has been used to moderate each wave of HR shortages (Shostak, 1987).

The general tendency to be expected in the future is therefore that labour-saving new technology will allow for fewer controllers handling greater amounts of air traffic.

In addition to this, and within a framework of more commercially oriented ATS, it can be expected that more contingency planning, downsizing and re-engineering will take place with the aim of eliminating controller surpluses where they exist and to avoid shortfalls.

13.4.1 Manpower Planning and Staffing Practices

As a consequence, new approaches in an enhanced MP and staffing system have to be developed and implemented which then can contribute to a higher efficiency and flexibility of the workforce.

Example: Gains are expected from a shift from team to individual rostering of controllers. The higher flexibility necessary to man operational positions more in accordance with traffic demands, would require for powerful rostering tools. These tools are currently under development in different States.

The ingenuity needed in industrial relations, collective bargaining and in solving the social and human factors problems associated with this undertaking will be tremendous.

Example: HR planners will need to ensure that future ATC systems and subsystems are well designed and that new procedures applied in staffing are efficient from a human factors standpoint.

The future approach in MP and management of HR has to be and will be more

- proactive,
- integrated,
- data-driven and
- flexible.

A necessary first step in this respect is to avoid a stop / start approach in recruitment and selection of *ab initio* trainees. This would help to avoid fluctuations and backlogs in subsequent training (Fisher and Marciano, 1996) and would lead to smoother to run recruitment and selection systems and earlier allocation of trainees to different ATC jobs.

13.4.2 Recruitment and Selection Strategies

For a long time now, candidates for *ab initio* training have been selected for "the controller job" mainly due to reactive MP.

If manpower requirements in different jobs could be identified before recruitment and selection takes place, target populations could be better addressed and attracted. It would then allow for the application of specific selection strategies (e.g. selection procedures for different jobs) and training programmes; both geared towards a specific job or even location.

The strategy not to suspend the hiring and training of controllers could also be beneficial if surplus controllers could be more easily contracted out to other countries within Europe.

Controller shortages that ATSP organisations were experiencing in many countries in the last 30 years will lead to seeking out non-traditional sources of supply such as cross-border recruitment, including ready entry ATCOs.

13.5 Future Psychometrics

Psychometric theory has never and probably never will completely explain success in training or later on the job itself.

In psychometric terms, a test is called a good predictor if it explains a significant percentage in the variance of, for example, training scores. However, from a practical point of view a test is a good predictor, if it is able to discriminate between "good" and "bad" applicants, i.e. between those who will succeed in training and those who will not.

A clearer understanding of the links between psychometric tests and performance in ATC training and job and the identification of ability profiles which lead to success is one of the major challenges in the future, including non-linear relationships between predictor and criterion.

Another even greater challenge lies in better training design. There is evidence that enhanced training procedures and methods in the framework of "success orientation" are powerful tools to achieve higher success rates.

14. SUMMARY OF GUIDELINES AND RECOMMENDATIONS

The STF has reached a stage at which it has specified and then analysed the following questions:

- What are important criteria and how should these criteria be applied to selection tools and methods for *ab initio* trainee controllers?
- What selection tools are currently available or emerging?
- Are there any reliable and valid test batteries for *ab initio* trainee controllers and what are the results of their further evaluation?

The STF defined a set of criteria and established a method for the evaluation of those tests / test batteries that were made available for the purpose of evaluation and met the basic preconditions in regard to reliability and validity. The results of this undertaking together with a basic evaluation of some tests are given in a report that partly answered the questions addressed above (see EATCHIP, 1997).

However, it could be felt that these results are somewhat disappointing. In regard to this, the STF wants to stress the point that no final or definite answer can be given and that the development of enhanced selection tools is and will be a challenge for the future e.g. FEAST.

The STF felt that solutions to the various and numerous problems addressed in this document are achievable

- by ongoing and continuous improvements,
- by communal efforts,
- by implementing the above-mentioned guidelines.

The STF felt however that the results may reassure those who fear that reliable and valid test batteries do not exist or that it is impossible to know how to arrive reliably at the most informed selection decisions.

The results obtained and the guidelines given will greatly increase the knowledge about the selection of *ab initio* trainee controllers and the ways it should be applied.

Knowing how to select in a better way is not always a comfortable process. But once enhanced methods and tests have been tried, tested and validated, they may be integrated into a current selection system with more confidence and justification.

Page intentionally left blank

15. ANOTHER APPROACH IN SELECTION, RECRUITMENT AND TRAINING

15.1 Approaches in Selection: A Summary

The majority of information available and work done in the context of this and the preceding selection Guidelines released has been associated with a certain general approach and philosophy to recruitment and selection of *ab initios* that is prevalent in Europe and the United States. This approach is based on the starting point that although a certain minimum educational standard is a prerequisite to participation and selection for the job of air traffic controller, educational attainment in itself is not considered as the sole basis for selection and recruiting candidates for controller training.

Consequently, the main basis for selecting the “most suitable” candidates that is those that have the best chance of success in training and will become competent and safe controllers is based on basic abilities, knowledge and skills. A specific range of cognitive abilities and personality characteristics are thought to be more important predictors of success in the job than pure intelligence or ability to achieve high academic goals. Efforts in improving selection methods are therefore concentrated on developing predictive tests and other assessment procedures to measure these skills, abilities and competencies as key components of the selection system as is described in the major parts of the selection Guidelines and Reference Material produced. In practice a diversity or variants of selection tests and methods exists which follow more or less the approach described in this document.

This chapter now describes a different alternative that is followed in France. Here, the "testing" of candidates to select the "most suitable" is based only on academic achievement and a heavily accented measurement and assessment of knowledge, ability and skills in educational and academic subjects. The capacity to pass a series of written and oral exams of a scientific nature plus English language and possibly other languages and some optional subjects is the starting point for admission to full academic training in Civil Aviation Engineering at and above University level in France. This philosophy of recruitment is internally consistent with the ensuing nature of the specialised training and prerequisite academic achievement required before trainees may begin on-the-job training as Air Traffic Controllers or start in other, related careers outside ATC or even in other Civil Aviation areas. More than 95% of the recruited students become fully qualified controllers. This is partly due to the flexibility of a 'Relocation Policy' adopted, for example, in France (see Sub-chapter 5.3.5). This policy allows students to switch the team and ask for a new assignment after counselling (see “Proceedings of the First EUROCONTROL Selection Seminar - Current and Required Future Selection Work and Methods in the ECAC Area” [EATMP, 1999a, p.139]).

The purpose of this technical supplement is to describe and explain the fundamentals and procedures of the recruitment, selection and training system

in force in France for Air Traffic Controllers so as to present an alternative approach to that adopted and described in this chapter and the preceding selection Guidelines.

15.2 The Recruitment, Selection and Training of *Ab Initio* Controllers in France

15.2.1 General Information

The “Direction de la Navigation Aérienne (DNA)” (Air Navigation Head Office) is a directorate in the “Direction Générale de l’Aviation Civile (DGAC)” (Civil Aviation Head Office).

ATCO’s employed in the DNA are divided into two categories:

- “Ingénieur du Contrôle de la Navigation Aérienne (ICNA)” and
- “Technicien Supérieur d’Etudes et d’Expoitation de l’Aviation Civile (TSEEAC)”.

ICNA’s work mainly in Approach and En-Route Control Centres and TSEEAC’s work mainly in Tower Control. The current number of ICNA’s is some 4000 and of TSEEAC of the order of 1500. Only 200 of the TSEEAC are controllers. The current yearly recruitment requirement is about 200 for ICNA’s and 85 for TSEEAC’s.

ATCO’s of both categories are selected for, recruited and trained at the “École Nationale de l’Aviation Civile (ENAC)” (National School for Civil Aviation, Toulouse). Once candidates are recruited, they become national officials with the full status of civil servant in the DNA.

ENAC is also responsible for the selection, recruitment and training of technical staff as electronics engineers, ATS engineers and managers employed within DGAC and other profession in the aviation domain, including pilots.

15.2.2 Admission and Training at ENAC

ENAC has the status of a “Grande École” (the very best institutions of French higher education) and is specialised in civil aviation (see also Figure 3). To be considered as a Grande École, a school must be a member of the "Conference des Grandes Écoles", a governmental ruled institution and recognised by law. Grandes Écoles are not university institutions but characterised by their highly selective entrance requirements. They award high-level qualifications and form part of long courses of higher education (“enseignement supérieur long”). The distinction between a Grande École and a university refers solely to the institution and not the level of qualification. The level of qualification of a Grande École is equivalent to that of a university long course of higher education.

Thus, the admission, selection, recruitment and training including exams, etc., are laid out and controlled as governmental procedures that apply to all institutions at this level. To prepare for Grandes Écoles entrance competitions/exams, students must enrol in corresponding preparatory classes ("classes préparatoires"). Courses are held in selected public high schools and private institutions and run about two years.

The preparatory classes ("classes préparatoires aux Grandes Écoles") are of a literary, commercial or scientific nature. Highly selective, they prepare students for the competitive entrance examinations for the Grandes Écoles. Admission to the preparatory classes, the level of which corresponds to that of the first two years of a university, is itself open only to students with outstanding secondary school records and recommendations.

The information on the admission, examination and selection system below is based on the largest recruitment needs - that for ICNA's. However, the information is exemplar to other admission, examination, etc., procedures of other professional categories outlined above. The information is based on information contained on a dedicated Web site of ENAC (www.enac.fr in French language) and relates to the admission year 2002.

15.2.3 Admission to ICNA Exam

There are a number of preconditions for an individual to be admitted to the exam procedure for selection. Notably a maximum age limit of 26 years is set and candidates must have the French nationality.

Admission to the entrance examination at ENAC is done via the means of inscription as a (potential) student. Initial Registration is done by electronic means using an identifying password.

All candidates will receive by the end of March a computerised acknowledgement of inscription based on the information they have furnished.

To be admitted candidates must either:

- have completed the second year of a preparatory class (or the second year of an integrated preparatory class) at Technical Colleges (see Figure 3) or
- already possess (before 1 November of the year of admission) a scientific or technology-based diploma achieved at these Technical Colleges.

Documentary proof of the above pre-conditions should be presented at the time of the medical examination when the entry file is being prepared.

15.2.4 Examination

The examination of candidates is done at fixed dates and times in dedicated cities throughout France by means of written and oral exams. The exam centres will be open according to the number of candidates applying.

Candidates are encouraged to at least choose two centres in order of preference to be sure.

Table 5 gives the academic areas and the duration of these exams.

Table 5: Common written obligatory and optional exams

Exams	Duration
Maths[1]	4 hours
Physics[1]	4 hours
French	4 hours
English[1]	2 hours
Written obligatory option one choice	
Either Maths[1]	4 hours
Or Physics[1]	4 hours
Or Technology[1]	4 hours
Optional written exams one choice (bonus)	
Either Aeronautical Knowledge	2 hours
Or a second living language (German, Spanish, Italian or Russian)	2 hours

[1]These exams are in the form of Multiple Choice Questionnaires.

Only those candidates declared admissible after the written exams are invited to the oral exams (see Table 6).

Table 6: Obligatory oral exams

Exams	Duration
Maths	30 minutes
Physics	30 minutes
French	30 minutes
English	30 minutes

Only those candidates who have score a minimum mark on the sum of the written and oral exams are considered for admission.

15.2.5 The Content of the Exams

Mathematics

- (Obligatory) Algebra and Geometry: General Algebra, Linear Algebra and Geometry; Analysis and Differential Geometry: Suites and Functions, Differential Calculus and Integrals, Series, Differential Equations, Vector Functions, Differential Geometry.

- (Optional) Algebra and Geometry: In addition to above: Numbers and Structures, Fourier Series, Functions with several Variables.

Physics

- (Obligatory) Newtonian Mechanics, Thermodynamics, Electrostatics, Magnetostatics, Electromagnetic, Electrocinetics, Optical Geometry, Optical Undulation.

- (Optional, in addition to the above) Mechanics of Fluid, Thermal Diffusion.

French

- (Written) A summary or commentary on a given text to evaluate the capacity of the candidate to present ideas clearly and logically in correct French.

- (Oral) Analysis and comment on a given text of a general nature in front of several examiners of French.

English

- (Written) Questions on subjects of everyday life or general news. This exam aims to measure knowledge of vocabulary of language structure.

- (Oral) The exam consists of a conversation with an examiner after listening to a recording.

Technology

- (Obligatory) Mechanics, Automation, Systems Studies.

Second Language

- The objective is to judge the breadth of vocabulary and knowledge of grammar as well as the ease with which a candidate can translate a text. This exam can comprise a version on a non-technical subject, some general themes or the report of a short report.

Aeronautical Knowledge

- Air Traffic, ANS, Navigation, Meteorology, Aerodynamics and Technical Aeronautics.

15.2.6 Training and Assignment as ICNA

Training of an "Ingénieur du Contrôle de la Navigation Aérienne" (air traffic control engineer) is given at ENAC and lasts three years.

Recruits will receive during eighteen months at the ENAC basic training giving the necessary technical knowledge in the field of aeronautics. In addition, the programme covers the management and economics of air transport, the organisation of French and international civil aviation, etc. This is a high level training which is open minded including training in the English language.

During training there are a number of training periods (stages). The stages comprise:

- a six-week training period in an anglophone country,
- a private pilots licence training,
- a training period in an aviation company.

The initial academic training period of three years is completed by the writing and presentation of a "mémoire", and is followed by a period of up to two years at an Air Traffic Control centre. Here students will work towards the acquisition of a controller qualification which will allow them to fully exercise their responsibilities.

15.2.7 Conditions of Training

Training is remunerated with an automatic pay increase during each of the three years. In return, recruits will undertake to stay seven years in public service at the end of the training.

The integration of controller engineers (Conversion) to ENAC is done in intakes of 32 pupils in order of performance in the exams.

ENAC has a campus with residential and leisure facilities. Halfway through the training, students will have to choose according to their training results the centre of assignment.

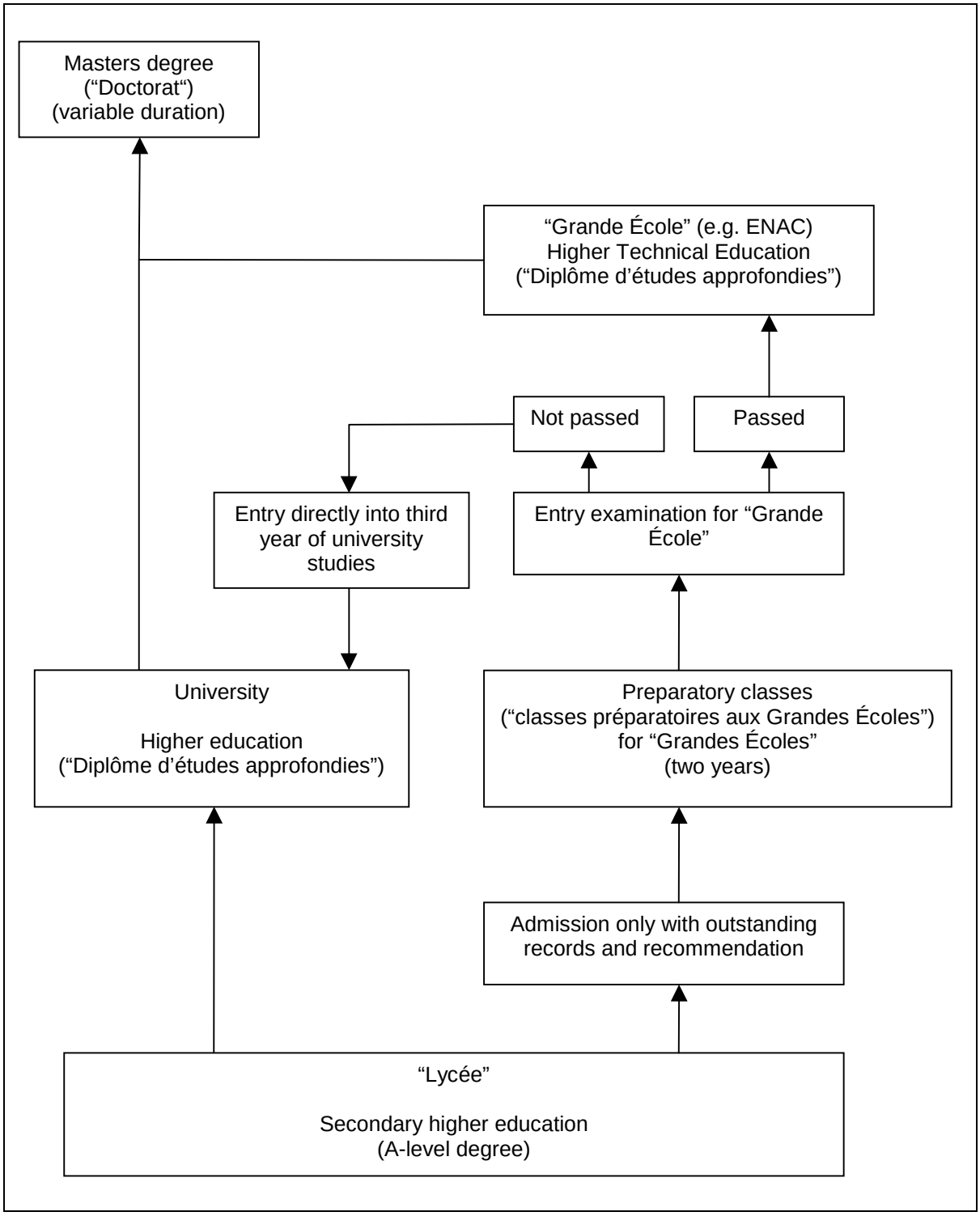

Figure 3: The French Higher Educational System

Page intentionally left blank

REFERENCES

Ackerman, P. L. and Kanfer, R. (1993). Integrating laboratory and field study for improving selection: development of a battery for predicting Air Traffic Controller success, *Journal of Applied Psychology, 78, (3),* 413-432.

Anastasi, A. (1982). *Psychological Testing*, (5th edition), New York: MacMillan Publishing Co., Inc.

Anastasi, A. (1988). *Psychological Testing*, (6th edition), New York: MacMillan Publishing Co., Inc.

Berg, R. H., Hvilsted, S. and Ramanujam, P. S. (1996). Peptide oligomers for holographic data storage, *Nature*, *383,* 505-508.

Brehmer, B. (1996). *What does the situational interview measure?* MRU Report No. 22. Norrköping: Luftfartsverket.

Brokaw, L. D. (1984). Early research on controller selection 1941-1963. In: S. B. Sells, J. T. Dailey and E. W. Pickrel (Eds.), *Selection of air traffic controllers*. Report No. FAA-AM-84-2. Washington DC: FAA Office of Aviation Medicine.

Brooker, P. (1996). Air Traffic Control, *Newsline Daily Lockheed Martin International S.A*., 04 Sep 1996, 2-4.

CAST Consortium (1999a). *WP1: Current and future ATM systems*. (NLR-TR-98233). Amsterdam: NLR.

CAST Consortium (1999b). *WP2: Future ATCO description*. (NLR-TR-98496). Amsterdam: NLR.

CAST Consortium (1999c). *WP3: Future ATCO selection and training. Volume II - Future ATCO selection*. (NLR-TR-99315-VOL-2). Amsterdam: NLR.

Cattell, R. B. (1973). *Personality and Mood by Questionnaire*. San Francisco: Jossey-Bass.

Cattell, R. B. and Warburton, F. W. (1967). *Objective Personality and Motivation Tests*. Urbana: University of Illinois Press.

Committee to Develop Standards for Educational and Psychological Testing (CDSEPT) (1999) of The American Educational Research Association, The American Psychological Association, and The National Council on Measurement in Education, co-chaired by E. Baker and P. Sackett. *The Standards for Educational and*

Psychological Testing. Washington, DC: American Educational Research Association.

Crocker, L. and Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston.

Cronbach, L. J. (1990). *Essentials of Psychological Testing* (5th edition), New York: Harper Collins Publishers, Inc.

Downing, E., Hesselink, L., Ralston, J. and Macfarlane, R. (1996). A three-colour, solid-state, three-dimensional display, *Science*, *273,* 1185-1189.

EATCHIP Human Resources Team. (1996). *Report on Issues in ATCO Manpower Planning.* HUM.ET1.ST03.1000-REP-01. Ed. 1.0. Brussels: EUROCONTROL.

EATCHIP Human Resources Team. (1997). *Information on Available and Emerging Selection Tests and Methods for Ab Initio Trainee Controllers.* HUM.ET1.STO4.1000-REP-01. Ed. 1.0. Brussels: EUROCONTROL.

EATCHIP Human Resources Team. (1998a). *Selection Tests, Interviews and Assessment Centres for Ab Initio Trainee Controllers: Technical Supplement.* HUM.ET1.STO4.1000-GUI-03-02. Ed. 1.0. Brussels: EUROCONTROL.

EATCHIP Human Resources Team. (1998b). *A Framework for Applying Cost-Benefit Considerations in the Recruitment and Selection Process for Ab Initio Trainee Controllers.* HUM.ET1.ST11.4000-REP-01. Ed. 1.0. Brussels: EUROCONTROL.

EATCHIP Human Resources Team. (1998c). *ATS Manpower Planning in Practice: Introduction to a Qualitative and Quantitative Staffing Methodology*. HUM.ET1.ST02.2000-REP-01. Ed. 1.0. Brussels: EUROCONTROL.

EATCHIP (1998d). Human-Machine interface Catalogue for EATCHIP Phase III - General Principles. DPSET1.ST07. Brussels: EUROCONTROL.

EATMP Human Resources Team. (1999a). *Proceedings of the First EUROCONTROL Selection Seminar – Current and Required Future Selection Work and Methods in the ECAC Area.* HUM.ET1.ST04.1000.REP-02. Ed. 1.0. Brussels: EUROCONTROL.

EATMP Human Resources Team. (1999b). *Air Traffic Controller Training at Operational Units.* HUM.ET1.ST05.4000-GUI-01. Ed. 2.0. Brussels: EUROCONTROL.

EATMP Human Resources Team (2000a). *Update of Information on Selection Tools and Methods for Ab Initio Trainee Controllers.* HRS/MSP-002-REP-01. Ed. 1.0. Brussels: EUROCONTROL.

EATMP (2000b.) *Human Resources Programme - Stage 1: Programme Management Plan*. Brussels: EUROCONTROL.

EATMP Human Resources Team (2000c*). Guidelines for Common Core Content and Training Objectives for Air Traffic Controllers Training (Phase II)*. HUM.ET1.ST05.1000.GUI-02. Ed. 1.0. Brussels: EUROCONTROL.

EATMP Human Resources Team (2000d). *Marketing and Communication of the ATCO Job - Phase 1: Study.* HRS/MSP-002-REP-02. Ed. 1.0. Brussels: EUROCONTROL.

EATMP Human Resources Team (2000e). *Guidelines for ATCO Manpower Planning Processes.* HUM.ET1.ST03-1000-GUI-02. Ed. 1.0. Brussels: EUROCONTROL.

EATMP Human Resources Team (2000f). *European Manual of Personnel Licensing – Air Traffic Controllers.* HUM.ET1.ST08.10000-STD-01. Ed. 1.0. Brussels: EUROCONTROL.

EATMP (2001a) *European Convergence and Implementation Plan for the Years 2002 to 2006 – Level 2.* Edition 1.1. Brussels: EUROCONTROL.

EATMP Human Resources Team. (2001b). *Guidelines for Selection Procedures and Tests for Ab Initio Trainee Controllers (Revised).* HRS/MSP-002-GUI-01. Ed. 2.0. Brussels: EUROCONTROL.

EATMP Human Resources Team. (2001c). *Guidelines for Common Core Content and Training Objectives for Air Traffic Controllers Training (Phase I: Revised)*. HRS/TSP-002-GUI-01. Ed. 2.0. Brussels: EUROCONTROL.

EATMP Human Resources Team. (2001d). *Characteristics of Recruitment and Pre-selection of Ab Initio Trainee Controllers (Revised).* HRS/MSP-002-GUI-02. Ed. 2.0. Brussels: EUROCONTROL.

EATMP Human Resources Team (2001e). *Long-term ATCO Manpower Planning Simulation (LAMPS) - User Guide.* HRS/MSP-003-REP-01. Ed. 1.0. Brussels: EUROCONTROL.

Fisher, E. G. and Marciano, V. (1996). Managing Human Resources Shortages in a Unionized Setting: Best Practices in ATC.

Fleming, P. (2001). *Future Air Traffic Management systems and their impact on the selection of en route air traffic controllers for the Masstricht*

Upper Area Control Centre, a dissertation work towards MSc in Air Transport Mangement.

Garcia-Avello, C. and Swierstra, S. (1997). *Human role in ATM: Support for decision making*, Paper presented at AGARD Workshop of the Mission Systems Panel on "Support for Decision Making: Optimisation - Automation", Budapest, 27-29 May 1997, Brussels: EUROCONTROL.

Gatewood, R. D. and Feild, H. S. (1987). *Human Resource Selection*, New York: CSB College Publishing.

Ghiselli, E. E. (1966). The validity of a personnel interview, *Personnel Psychology, 19,* 389-395.

Haglund, R., Backman, B., Bengtsson, J., Brehmer, B. and Hoffman, P-I. (1996). *Principles of Institutional Training for ATCO Students in TWR and APP Control*. MRU Report No. 26. Norrköping: Luftfartsverket.

Hambleton, R. K. and Swaminathan, H. (1985). *Item response theory: Principles and applications,* 147-200. Hingham, MA: Kluwar.

Hansen, J. C. (1990). Interest Inventories, In: G. Goldstein, and M. Hersen, (Eds.), *Handbook of Psychological Assessment*, (2nd Edition), 148-172. New York: Pergamon Press.

Hardeman, T. (1994). Siemens cardian guards the ground, *Aerospace (Asia Pacific), 58,* February 1994.

Heron, R. M. (1990). Use of Magnetoencephalography (MEG) and Electroencephalography (EEG) in processes of selection and training of air traffic controllers, in: *Proceedings of the ICAO Human Factors Seminar, Human Factors No. 4,* Circular 229-AN/137, 435-447.

Heron R. M., Weinberg H., Robertson, A. and Jantzen. K. (1995). Development of a multitasking device for measurement of air traffic controller skills, In: R. S. Jensen and L. A. Rakovan (Eds.), *Proceedings of the 8th International Symposium on Aviation Psychology*, April 24-27 1995, Ohio: The Ohio State University, 348-353.

Holland, J. P. (1985). *The Holland Vocational Preference Inventory (Revised).* Odessa: Psychological Assessment Resources.

Hopkin, V. D. (1995). *Human factors in Air Traffic Control*. London: Taylor and Francis.

Hunter, J. E. and Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96 (1),* 72-98.

ICAO. (1988). *International Standards and Recommended Practices Personnel Licensing,* Annex 1 to the convention on international civil aviation, 8th edition, July 1988, p. 37.

Kline, P. (1995). *The Handbook of Psychological Testing*, London: Routledge.

Landy, F. J. (1989). *Psychology of work behaviour*, (4th edition). Belmont, California: Brooks / Cole Publishing Company.

Landy, F. J. (1993). Job Analysis and Job Evaluation: The Respondent's Perspective. In: H. Schuler, J. L. Farr and M. Smith (Eds.), *Personnel Selection and Assessment*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behaviour and Human Performance, 3,* 157-189.

Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika, 39,* 247-264.

Manning, C. A. (1991). Individual differences in air traffic control specialist training performance. *Journal of the Washington Academy of Sciences, 81 (2),* 101-109.

Proulx, P. (1996). Meeting global ATC challenges together, *Journal of ATC*, Jul.-Sep. 1996, 76-77.

Reeves, D. J. and Booth, R. F. (1979). Expressed versus inventoried interests as predictions of paramedical effectiveness. *Journal of Vocational Behaviour, 15*, 155-163.

Reschly, D. J. (1990). Aptitude tests in educational classification and placement, In: G. Goldstein, and M. Hersen, (Eds.), *Handbook of Psychological Assessment*, (2nd Edition), 148-172. New York: Pergamon Press.

Schmidt, F. L. and Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings, *Psychological Bulletin, 124,* 262-274.

Schmidt, F. L., Mack, M. J. and Hunter, J. E. (1984). Selection utility in the occupation of U.S. park ranger for three modes of test use. *Journal of Applied Psychology, 69,* 490-497.

Schuler, R. S. and Huber, V. L. (1993). *Personnel and human resource management*, (5th edition). St. Paul: Minnesota: West Publishing.

Shostak, A. B. (1987). Technology, ATC and Labor-Management Relations, In: D. B. Cornfield (Ed.), *Workers, Managers and Technological Cgange*. New York: Plenum Press, 153-172.

Society for Industrial and Organisational Psychology (SIOP), Inc. (1987). *Principles for the validation and use of personnel selection procedures* (3rd ed.). College Park, MD: Author.

Stix, G. (1994). Aging Airways, *Scientific American*, *270,* May 1994, 96-104.

Subkoviak, M. J. (1984). Estimating the reliability of mastery-nonmastery classifications. In: R. A. (Ed.), *A guide to criterion-referenced test construction,* 267-291. Baltimore: Johns Hopkins University Press.

Suen, H. K. (1990). *Principles of Test Theories*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Task Force on Assessment Centre Standards (TFACS), (1980). J. Moses, (Chair) Standards and ethical considerations for Assessment Centre operations. *The Personnel Administrator*, *25,* 35-38.

Tenopyr, M. L. (1987). Testing. In: R. L. Craig (Ed.), *Training and Development Handbook*, (3rd Edition), 286-300, New York: McGraw-Hill Book Company.

The Economist, (1996). Hi-fi technology - ultraspeakers, August 24th 1996, 66-68.

Vandenberghe, C., Nicolas, B. and Nugent M. (2001). *Air Traffic Statistics and Forecasts*, Number of Flights by Region 1974 –2008, 2015 (Total Airspace), Report No. 01.70.03, dated July 2001. Brussels: EUROCONTROL.

Weinberg, H. (1995). *The use of EEG and multitasking to determine transfer of training factors for ATC students and selection criteria for ATC applicants*, (Transport Canada Technical Report). Montréal: Transportation Development Centre.

Wiggins, J. S. (1973). *Personality and Prediction, Principles of Personality Assessment*, Menlo Park, CA: Addison-Wesley Publishing Company.

FURTHER READING

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*. New York: John Wiley and Sons.

Binning, J. F. and Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology, 74 (3),* 478-494.

Burke, E. F. and Hunter, D. R. (1993). Meta-analysis of aircraft pilot selection procedures, in: R. S. Jensen and D. Neumeister, (Eds.), *Proceedings of the seventh international symposium on aviation psychology*, *Vol. 1,* 413-417, April 26-29, 1993. Ohio: The Ohio State University.

Buros, O. K. (Ed.), (1978). *The VIIIth Mental Measurement Yearbook*, New Jersey: Highland Park: Gryphon Press.

Butterworth-Hayes, P. (1996). Costs of CNS/ATM underestimated, *Jane's Airport Review*, (October 1996), 27-29.

Cascio, W. F., Alexander, R. A. and Barrett, G. V. (1988). Setting cut-off scores: legal, psychometric, and professional issues and guidelines. *Personnel Psychology, 41*, 1-26.

Cattell, R. B., and Johnson, R. C. (Eds.), (1986). *Functional Psychological Testing,* New York: Brunner Mazel.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, *19*, 15-18.

Cross, M. (1990). *Changing Job Structures*. Oxford: Heinemann Newnes.

Diener, G. and Skoniezki, A. (1996). Overview from the Joint Chairmen, *ATC Training and Simulation Conference*, Langen, 29-31 October 1996.

Dillon, W. R. and Goldstein, M. (1984). *Multivariate analysis: Methods and applications*. New York: John Wiley and Sons.

Eissfeldt, H., Broach, D. and Heil, M. (Eds.), (2002). *Staffing the ATM System – The Selection of Air Traffic Controllers, Aldershot*. Ashgate (in press).

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*, 327-358.

Hocking, R. (1976). The analysis and selection of variables in linear regression. *Biometrics, 32,* 1-49.

Latham, G. P. and Finnegan, B. J. (1993). Perceived practicality of unstructured, patterned, and situational interviews. In: H. Schuler, J. L. Farr, and M. Smith, (Eds.), *Personnel Selection and Assessment,* 41-55. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Latham, G. P. and Saari, L. M. (1984). Do people do what they say? Further studies on the situational interview*. Journal of Applied Psychology*, *69,* 569-573.

Latham, G. P., Saari, L. M., Pursell, E. D., and Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, *65 (4),* 422-427.

Maurer, S. D. and Fay, C. (1988). Effect of situational interviews, conventional structured interviews, and training on interview rating agreement: An experimental analysis. *Personnel Psychology, 41,* 329-344.

Norusis, M. J. (1993). *SPSS Base System*. Release 6.0 Chicago: SPSS Inc., 311-376.

Pascoe, R. (1996). Future hidden costs, *Jane's Airport Review*, October 1996, 30-32.

Schmidt, F. L., Hunter, J. E. and Urry, V. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology*, *61,* 478-485.

Schuler, H. (1993). *Lehrbuch Organisationspsychologie*, Bern: Verlag Hans Huber. 270-273.

Sparrow, P. and Hiltrop, J. M. (1994). *European human resources management in transition*, 1st edition. Hempstead, Hertfordshire: Prentice Hall International (UK) Ltd.

Stevens, E. H. and Busse, D. E. (1996). Integration of ADS / CPDLC into air traffic automation systems for the Asia/Pacific region*, Journal of ATC*, Jul.-Sep. 1996, 42-47.

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. New York: John Wiley and Sons.

Vidler, N. (1996). Human factors aspects in CNS/ATM systems, *Journal of ATC*, Jul.-Sep. 1996, 72-75.

Winer, B. J., Brown, D. R. and Michels, K. M. (1991). *Statistical principles in experimental design*. (3rd Edition). New York: McGraw-Hill.

GLOSSARY OF TERMS

For the purpose of this document the following definitions shall apply:

Ability[4]: What a person brings to the job situation without specialised, job specific training, education or experience. There are many kinds of abilities. As regards ATCO selection, the focus is on relatively broad mental abilities.

Ability Test: A standardised and controlled method for measuring abilities in different areas (e.g. memory, spatial, perception and attention). Ability test covers speed or power tests, paper-and-pencil or computer-based tests. Ability tests are developed in order to predict future performance in areas in which individuals are not currently trained. Ability tests used in selection of ATCO candidates cover a more or less broad range of abilities pertinent to the job of an ATCO.

***Ab Initio* Trainee Controllers**: Selected individuals, with no previous relevant qualifications, who are given basic instruction and training to enable them to obtain theoretical qualifications.

Airborne Collision Avoidance Systems (ACAS): A way in which aircraft themselves can be protected by the use of short-term conflict alert, whereas ground computers process radar and flight plan information to detect possible conflicts.

Automatic Dependent Surveillance (ADS): ADS is a function in which aircraft automatically transmit, via a datalink, data derived from on-board navigation systems. This data is displayed to the controller on a screen that resembles a radar screen (Vidler, 1996).

Affirmative Action: The use of racial, ethnic or linguistic preferences in recruitment to jobs, hiring, admissions and contracting. This positive discrimination favours those who tend to (or could) suffer from discrimination.

Air Traffic: All aircraft in flight or operating on the manoeuvring area of an aerodrome (ICAO Doc 9569 Definitions).

Air Traffic Service (ATS): A generic term meaning variously, flight information service, alerting service, air traffic advisory service (area control service, approach control service or aerodrome control service). (ICAO Doc 9569 Definitions).

Applicant: A person who applies for an *ab initio* training vacancy.

Application Form: Form to be filled in when applying for *ab initio* training.

[4] There is yet no final position in the professional literature with regard to the question whether abilities are ‘innate’ qualities or have been already partly shaped by general education and experience.

Assessment Centre (AC): A behaviour-oriented method of selection to assess the social and action competence of applicants in job-related situations. Applicants are observed and assessed by trained observers in various standard situations (e.g. role-play, teamwork) with regard to certain competences.

Battery: A set of tests standardised on the same population, so that norm-referenced scores on the several tests can be compared or used in combination for decision-making (CDSEPT, 1999).

Candidates: Individuals whose applications have been accepted for further consideration in the selection process but have not yet been selected finally as *ab initio* trainee controllers.

Communication, Navigation and Surveillance (**CNS) / ATM**: The way in which enhanced capabilities of satellite-based navigation and digital datalink communication systems will permit the next generation of ATM systems to combine the features of ADS and Controller-to-Pilot Datalink Capability (CPDLC) with the conventional ATC functions. The combination of these enabling CNS technologies and their application to ATM has become known as CNS/ATM.

Composite Score: A score that combines several scores by a specified formula (CDSEPT, 1999).

Concurrent Validity: A demonstrated relationship between job performance and scores on selection instruments obtained at approximately the same time (SIOP, 1987).

Construct-related Validity: The process of identifying psychological construct domains that overlap significantly with the performance domain and then developing predictors that adequately sample these construct domain.

Content-related Validity: The extent to which predictors directly sample the performance.

Controllers in:

- **Area Control Centre (ACC) or Upper Area Control Centre (UAC) -** ATCOs who are responsible for ATS on the en route phases of flight. They can use radar or other means.

- **Terminal Control Area (TMA)** - ATCOs who are responsible for ATS in a transition phase between en route and aerodrome / approach control. Terminal area normally applies in the vicinity of major airports and/or groups of airports.

- **Approach** (with or without radar) - ATCOs who are responsible for the arrival and departure phases normally in the vicinity of 40 miles radius of an airport up to a defined altitude.

- **Tower** - ATCOs who are responsible for movements in the immediate vicinity of an airport, normally five-ten miles, including control of aircraft as they move on the ground (ground movements) and as they take off or land up to a defined altitude.

Correlation: The degree to which two sets of measurements vary together (SIOP, 1987).

Criterion: A measure of job performance or job behaviour, such as training score or supervisory ratings.

Criterion-related Validity: The extent to which the operational measure of behaviours in the performance domain (i.e. the operational criterion measure) is related empirically with the developed predictor.

Critical Incidents Technique (CIT): A critical incident is a description of an event that was central (or critical) to success or failure on a particular task (Landy, 1989). CIT is a qualitative approach to obtain descriptions of specific, observable behaviour. The CIT also gives descriptions of the context (situation) in which the behaviour occurred and the consequence (outcome) of the behaviour described.

Cross Validation: The application of a scoring system or set of weights empirically derived in one sample to a different sample from the same population to investigate the stability of relationships based on the original weights (SIOP, 1987).

Customer: Any person or any unit receiving a service or product.

Cut-off Score: A score in a predictor distribution of test scores below which candidates are rejected.

Decision-making Ability: The mental process by which operators recognise, analyse and evaluate information about themselves, the air traffic and the operational environment, leading to a decision.

Decision-making of Selectors: The process of accurate separation between ‘valid positive’ and ‘valid negative’ applicants by the use of valid information from tests and other selection tools by decision makers.

Dependent Variable: The response, or the variable which is hypothesised to be affected by the independent variable.

Discriminant Analysis: The procedure of classifying the individuals into two different groups on the basis of a test.

Face Validity: The extent to which a test appears to measure what it was supposed to measure.

Fairness: A social, not psychometric, concept; a joint function of the procedure, the job, the population and how the scores are used (SIOP, 1987).

Feasible: Capable of being done successfully given practical constraints (i.e. economically practical and technically possible) without misleading or uninterpretable results (SIOP, 1987).

Initial Training: The training phase in ATC Training that includes technical subjects, ATC theory and practice including practical training conducted on simulators. The object of initial training is to prepare a trainee for pre-On-the-Job Training (pre-OJT) at an ATC unit. Initial training might be divided into two phases:

(i) **Basic Training**: Training designed to impart fundamental knowledge and skills, to enable student ATCOs to progress to specialised ATC training.

(ii) **Rating Training**: (training in the rating discipline) Specialised ATC training to provide knowledge and skills related to a job category and appropriate to the discipline to be pursued in the ATS environment.

Interval Scale: A scale in which the difference between scale points at all points of the scale is equal (Kline, 1995).

Interview: A structured, or unstructured, on one-to-one or one-to-several basis conversation with a job applicant for the express purpose of identifying knowledge, skills, abilities and behaviours that may be predictive of success in subsequent training and in the job. To this end biographical, situational, and performance- and/or motivation-related information from the applicant is taken into consideration, in order to make a decision on employment.

Item Analysis: The process of assessing certain characteristics of test items, usually the difficulty value, the discriminating power, and sometimes, the correlation with an external criterion (CDSEPT, 1999).

In item analysis each item of a test is correlated with the total score of the test. Since the aim of the analysis is to produce a homogeneous test, each item should be measuring what the test measures. It is further analysed, if appropriate, whether the items of a test have sufficient selectivity, i.e. the items are able to discriminate between low and high performances and are at the required difficulty level.

Item Difficulty: An index that indicates the percentage of candidates who answered that item correctly.

Job: A group of a certain number of core tasks, direct support tasks and indirect support tasks which require a certain level of abilities, knowledge and skills.

Job Analysis: The systematic collection of data describing the tasks that comprise a job and the knowledge, skills, abilities and other characteristics that enable an individual to carry out those tasks (Landy, 1993).

Job Description: A high level overview of tasks which form a specific job and their required level of abilities, knowledge and skills.

Knowledge: The job-specific content or information gained through training, education and/or experience. Knowledge is built upon the foundation of mental abilities that a person brings to the situation.

Manpower Planning: A planning procees to ensure that the appropriate number of qualified controllers and other staff categories is available at the right time in all work locations and their associate functions in order to match the changing requirements of ATM and to cover operational requirements.

Mental Abilities: The operations of the mind. They are the fundamental, basic, and to some degree abstract, capacities and processes of perceiving, thinking, and deciding.

Multiple Regression: A technique in which a correlation is obtained between a number of variables and a criterion (Kline, 1995).

Norm-referenced Test: An instrument for which interpretation is based on the comparison of a test taker's performance to the performance of other people in a specified group (CDSEPT, 1999).

Norms: Statistics or tabular data that summarise the test performance of specified groups (e.g. test takers throughout the country) (SIOP, 1987).

A norm is a standard of achievement in tests or questionnaires usually derived from the average achievement of a large group of individuals. The interpretation of results in tests or questionnaires is normally based on a comparison of the individuals response with those previously obtained to establish the appropriate standard for the test scores.

Numerical Ability: The ability to manipulate numbers accurately and rapidly.

On-the-Job Training (OJT): The integration in practice of previously acquired job-related routines and skills under the supervision of a qualified coach in a live traffic situation ("Air Traffic Controller Training at Operational Units" [EATMP, 1999b]).

The training enables student controllers to check out as operational controllers at a specific operational unit.

Operational Controller: The holder of a certified qualification to permit that individual to control air traffic at a specific unit.

Ordinal Scale: A scale in which it is possible to rank or order all categories according to some criterion.

Paper Sift: The sift of application forms and attached application material in order to pre-select candidates according to previously defined criteria.

Performance: The effectiveness and value of work behaviour and its outcomes (SIOP, 1987).

Percentile: A value above and below which a specified percentage of cases fall.

Percentile Rank: The percentage of scores in a specified distribution that fall at or below the point at which a given score lies (CDSEPT, 1999).

Personality Tests: Questionnaires (or inventories) to measure personality traits or behavioural tendencies (CDSEPT, 1999).

Power: The probability that a statistical test will yield statistically significant results (SIOP, 1987).

Predictive Validity: A demonstrated relationship between test scores of applicants and some future behaviour on the job (SIOP, 1987).

Predictor: A measure used to predict criterion performance, for example, scores on a test, or judgements of interviewers (SIOP, 1987).

Pre-On-the-Job Training (Pre-OJT): The training phase in ATC Training of locally based training during which extensive use of simulation using site-specific facilities will enhance the development of previously acquired routines and abilities to an exceptionally high level of achievement.

Psychological Testing: The measurement of some aspects of human abilities by standardised procedures consisting of carefully prescribed content, methods of administration, and interpretation. Tests are usually administered by qualified psychologists or specially trained staff according to professional and ethical principles.

Psychometric: Pertaining to the measurement of psychological characteristics such as abilities, personality traits, achievement, skill and knowledge (CDSEPT, 1999).

Psychomotor Test: A physical test, which aims to measure human capabilities (such as coordination, steadiness, speed, etc.) of movements.

Quality Assurance: All planned and systematic measures required to create the necessary confidence in the ability of a product / service to meet agreed quality requirements or standards.

Raw Score: The obtained score on a measure without any transformation, such as percentages.

Reaction Time: Speed of reaction to a single stimulus.

Regression Equation: An algebraic equation used to predict criterion performance from predictor scores (SIOP, 1987).

Reliability: A measure of consistency or dependability or repeatability. The results of a particular testing session should be replicated if the same individuals were re-tested under similar conditions. Reliability of a test means

that the test behaves the same way under a variety of circumstances (e.g. the test yields similar results when different persons administer it, when the test is administered on different occasions, when alternative forms of the test are used, etc.).

Restriction of Range: A situation in which the **variability** of data in a sample is less than the variability in the population from which the **sample** has been drawn (SIOP, 1987).

Sample: The individuals who are actually tested from among those in the population to which the procedure is to be applied (SIOP, 1987).

Satellite Navigation and Global Positioning System (GPS): A way to allow aircraft not only to separate themselves by means of computerised on-board technology, but also to find their own optimum most direct route to their destination without advice from ATC by the change of ground-based navigational aids for air traffic to satellite navigation (Diener and Skoniezki, 1996).

Score: Any specific number resulting from the assessment of an individual (CDSEPT, 1999).

Sector: A means of dividing a piece of airspace for which controllers are responsible in order to make the workload more manageable.

Selection Instrument: Any method or device used to evaluate characteristics of persons (SIOP, 1987).

Selection Procedures: Selection procedures provide guidelines for accepting, or rejecting, applicants for a job on the basis of data from application forms, psychological tests, personality inventories or interviews.

Selection Ratio: The percentage of those candidates considered who are selected.

Situational Awareness: The continuous extraction of environmental information, integration of this information with previous knowledge to form a coherent mental picture, and the use of that picture in directing further perception and anticipating future events.

Situational Interview (SI): A specific type of structured and standardised interview based on hypothetical situations.

Skill: Competence to perform the work required by the job (SIOP, 1987).

Skill refers usually to the combination of ability and knowledge after training and practice which is required to perform a specific job.

Spatial Ability: The capacity to visualise two or three-dimensional figures when their orientation is changed.

Speed Test: A test in which performance is measured by the number of tasks performed in a given time. Examples are tests of typing speed and reading speed. Also, a test scored for accuracy where the test taker works under time pressure (CDSEPT, 1999).

Speediness: The extent to which a test taker's score on a test depends on the rate at which work is performed rather than on the correctness of the response. One index of speediness is the percentage of test takers who do not complete the test (CDSEPT, 1999).

Stanine: An approximately normal transformation, where only nine score categories are allowed with a mean of 5 and standard deviation of 1.96.

Standard Error of Measurement: Standard deviation of errors of measurement, that is, differences between true and obtained scores (SIOP, 1987).

Student Air Traffic Controllers: Selected individuals who have successfully completed approved initial ATC instruction and training to enable them to obtain theoretical qualifications prior to commencement of Pre-OJT ("European Manual of Personnel Licensing – Air Traffic Controllers" [EATMP, 2000f]).

Supplier: Any person or any unit providing a service or a product.

Task: A piece of work, performed by an individual or individuals, which has a definite beginning and end, and results in a product or a service.

Task Analysis (TA): A group of analytical procedures directed at describing work activities in detail. A TA aims to provide a comprehensive breakdown of the content of tasks.

Team: A team in ATS is a group of two or more persons who interact dynamically and interdependently with specific assigned roles, functions and responsibilities.

Teamwork: Group effort applied to work.

Technical Manuals: Booklets prepared by test publishers to provide technical information on a test (CDSEPT, 1999).

Test: A standardised and controlled method for measuring abilities, knowledge or personality of applicants based on a sample of behaviour.

Test - Re-test Reliability: The stability of a test over time. This stability can be influenced due to changes in individuals, measurement errors, unclear test instructions, changes in the test administrators, guessing or the length of period of time between two testing sessions.

Training: The planned systematic development of the knowledge, understanding, skill, attitude and behaviour pattern required by an individual in order to perform adequately a given task or job.

Transitional Training: The stage following initial training during which site-specific theoretical knowledge and understanding will be transferred to the trainee using a variety of methods and during which skills will be developed through the use of site-specific simulations.

Unit Training: Training comprising transitional training, pre-OJT and OJT.

Utility: The relative value of an outcome with respect to a set of other possible outcomes (CDSEPT, 1999).

Validation: The effort through which the appropriateness and meaningfulness of interpretations from scores on a measure can be estimated (SIOP, 1987).

Validity: The degree to which inferences from scores are justified or supported by evidence (SIOP, 1987).

The extent to which predictor samples of behaviour effectively overlap with performance domains could serve as an operational definition of validity.

Validity Coefficient: A coefficient of correlation showing the strength of relationship in a criterion-related study (SIOP, 1987).

Validity Generalization: Applying validity evidence obtained in one or more situations to other similar situations on the basis of simultaneous estimation, meta analysis, or synthetic validation arguments (CDSEPT, 1999).

Variability: The spread or scatter of scores (SIOP, 1987).

Variable: A quantity that may take on any one of a specified set of values (SIOP, 1987).

Variance: A statistic which implies the sum of squares of deviations about the mean.

Page intentionally left blank

ABBREVIATIONS AND ACRONYMS

For the purposes of this document, the following abbreviations and acronyms shall apply:

16PF	Sixteen Personality Factor (questionnaire)
3-D	Three-Dimensional
AC	Assessment Centre
ACAS	Airborne Collision Avoidance System
ACC	Area Control Centre
ADS	Automatic Dependent Surveillance
AIS	Aeronautical Information Services
ANS	Air Navigation Services
ATC	Air Traffic Control
ATCA	ATC Assistant
ATCO	Air Traffic Controller / Air Traffic Control Officer *(US/UK)*
ATM	Air Traffic Management
ATS	Air Traffic Services
ATSP	Air Traffic Service Provider
BOI	Biographical-oriented Interview
CAST	Consequences of future ATM systems for air traffic controller Selection and Training
CBA	Cost-Benefit Analysis
CBT	Computer-Based Training
CDG	Core Drafting Group (Selection)
CD-ROM	Compact Disc Read-Only Memory
CDSEPT	Committee to Develop Standards for Educational and Psychological Testing

CFMU	Central Flow Management Unit *(EUROCONTROL)*
CIP	Convergence and Implementation Programme
CIT	Critical Incidents Technique
CNS	Communication, Navigation and Surveillance
CPDLC	Controller-to-Pilot DataLink Capability
CPI	Californian Psychological Inventory
CV	Curriculum Vitae
DAT	Differential Aptitude Test
DGAC	Direction Générale de l'Aviation Civile *(France)*
DIS	Director(ate) Infrastructure, ATC Systems and Support *(EUROCONTROL Headquarters; EATMP)*
DIS/HUM	Human Factors and Manpower Unit *(EUROCONTROL Headquarters; EATMP; also known as HUM (Unit))*
DNA	Direction de la Navigation Aérienne *(France)*
EATCHIP	European Air Traffic Control Harmonisation and Integration Programme *(now EATMP)*
EATMP	European Air Traffic Management Programme *(formerly EATCHIP)*
ECAC	European Civil Aviation Conference
ECIP	European Convergence and Implementation Plan
EEG	Electroencephalogram
ENAC	Ecole Nationale de l'Aviation Civile *(France)*
ET	Executive Task *(EATCHIP)*
EUROCONTROL	European Organisation for the Safety of Air Navigation
FDA	Flight Data Assistant
FEAST	First European ATCO Selection Test package
GPS	Global Positioning System
GUI	Guidelines *(EATCHIP/EATMP)*

HMI	Human-Machine Interface
HR	Human Resources
HRM	Human Resources Management
HRS	Human Resources Programme *(EATMP)*
HRT	Human Resources Team *(EATCHIP/EATMP)*
HUM	Human Resources (Domain) *(EATCHIP/EATMP)*
HUM (Unit)	See 'DIS/HUM'
IANS	Institute of Air Navigation Services *(EUROCONTROL, Luxembourg)*
ICAO	International Civil Aviation Organization
ICNA	Ingénieur du Contrôle de la Navigation Aérienne *(France)*
IFR	Instrument Flight Rules
ISO	International Standards Organisation
IVR	Immersive Virtual Reality
KASOs	Knowledge, Abilities, Skills and Other characteristics
MEG	Magnetoencephalography
MMPI	Minnesota Multiphasic Personality Inventory
MP	Manpower Planning
MSG	Manpower Sub-Group *(EATCHIP/ EATMP, HRT)*
MSP	Manpower Sub-Programme *(EATMP, HRS)*
OJT	On-the-Job Training
OJTI	On-the-Job Training Instructor
REP	Report
SDE	Senior Director, Principal EATMP Directorate *(EUROCONTROL Headquarters)*
SI	Situational Interview

SIOP	Society for Industrial and Organizational Psychology, Inc.
SME	Subject Matter Expert
SPSS	Statistical Package for the Social Sciences
ST	Specialist Task *(EATCHIP)*
STANINE	STAndard NINE score
STATFOR	Specialist Panel on Air Traffic Statistics and Forecast *(EUROCONTROL)*
STF	Selection Task Force
TA	Task Analysis
TFACS	Task Force on Assessment Centre Standards
TMA	Terminal Control Area
TORs	Terms Of Reference
TS	Technical Supplement
TSEEAC	Technicien Supérieur d'Etudes et d'Exploitation de l'Aviation Civile *(France)*
UAC	Upper Area Control Centre
VFR	Visual Flight Rules
VR	Virtual reality
WWW	World Wide Web

REVIEW GROUP (Selection Experts User Group)

NAME	ORGANISATION, COUNTRY
Chairman	
Mr. Hermann Rathje	EUROCONTROL Headquarters
Rapporteur	
Mr. Zvi Golany	EUROCONTROL Headquarters
Members	
Mr. Bertil Andersson	LFV, Sweden
Ms. Lucilla Barducci	ENAV, Italy
Ms. Lilian Biber-Klever	LVNL, The Netherlands
Mr. Ragnar Bjørkhaug	NATAM, Norway
Ms. Emmanuelle Blanc	AUSTROCONTROL, Austria
Mr. Fidel Chetcuti	EUROCONTROL Headquarters
Ms. Inez Laaser	DFS, Germany
Mr. Hinnerk Eissfeldt	DLR, Germany
Mr. Marian Epurescu	ROMATSA, Romania
Mr. Andrej Grebensek	CAA, Slovenia
Mr. Peter Hoffmann	AUSTROCONTROL, Austria
Mr. Andy Hutchinson	NATS, United Kingdom
Mr. Christian Langer	AUSTROCONTROL, Austria
Mr. Jørgen Lund Hansen	NAVIAIR, Denmark
Mr. Rami Malik	HUNGAROCONTROL, Hungary
Mr. Pat O'Doherty	EUROCONTROL IANS
Mr. Jean-Michel Pubellier	DNA, France
Ms. Anja Schwab	DFS, Germany
Mr. Ludwig Tomaschett	Skyguide, Switzerland
Ms. Nona Vosnesensky	ROMATSA, Romania

Page intentionally left blank

CONTRIBUTORS (Selection Task Force II)

NAME	ORGANISATION, COUNTRY
Chairman	
Mr. Hermann Rathje	EUROCONTROL Headquarters
Secretary / Rapporteur	
Mr. Zvi Golany	EUROCONTROL Headquarters
Members	
Ms. Lucilla Barducci	ENAV, Italy
Mr. Grzegorz Celitan	PATA, Poland
Mr. Fidel Chetcuti	EUROCONTROL Headquarters
Mr. Brian Considine	EUROCONTROL IANS
Mr. Stefan Cucu	ROMATSA, Romania
Ms. Inez Laaser	DFS, Germany
Mr. Adrian Enright	EUROCONTROL IANS
Mr. Andy Hutchinson	NATS, United Kingdom
Mr. William King	EUROCONTROL IANS
Mr. Per Knudsen	NAVIAIR, Denmark
Mr. Ferenc Kóródi	HUNGAROCONTROL, Hungary
Mr. Pat O'Doherty	EUROCONTROL IANS
Mr Jean-Paul Pietrera	DNA, France
Mr. Michel Pistre	EUROCONTROL IANS
Mr. Jean-Michel Pubellier	DNA, France
Mr. Szilard Szirmak	HUNGAROCONTROL, Hungary
Mr. Ludwig Tomaschett	Skyguide, Switzerland

Page intentionally left blank