

PROGRAMME FOR  
HARMONISED AIR TRAFFIC  
MANAGEMENT RESEARCH  
IN EUROCONTROL



**DOC 99-70-05**



## **Validation Project - Final Report**

PHARE/NATS/VAL-8/FR; 1.0



**EUROCONTROL**

EUROCONTROL  
96 rue de la Fusée  
B-1130 BRUXELLES

**Prepared by: R. Whitaker, NATS**

**Date: May, 1999**

## Copyright Statement

The information contained in this report is the property of the PHARE Participants\*. The report or parts thereof may be published and or reproduced on the condition that due acknowledgement of authorship is made by quoting the copyright statement below. The copyright and the foregoing condition on publication and reproduction shall extend to all media in which the information may be embodied.

The information contained in this document is provided on an "as-is" basis and the PHARE Participants shall provide no express or implied warranty of any kind and shall accept no liability whatsoever for or in connection with the use of the information contained in the document.

\* The PHARE Participants are:

- the EUROCONTROL Agency;
- the CENA (Centre d'études de la navigation aérienne);
- the STNA (Service technique de la navigation aérienne);
- the NLR (Nationaal Lucht- en Ruimtevaartlaboratorium);
- the RLD (Rijksluchtvaartdienst);
- the LVNL (Luchtverkeersleiding Nederland);
- the DLR (Deutsches Zentrum für Luft- und Raumfahrt);
- the DFS (Deutsche Flugsicherung GmbH);
- the UK CAA (Civil Aviation Authority);
- the NATS (National Air Traffic Services);
- the DERA (Defence Evaluation and Research Agency)

### Copyright statement:

The copyright in this report vests in the European Organisation for the Safety of Air Navigation (EUROCONTROL); the CENA (Centre d'études de la navigation aérienne); the STNA (Service technique de la navigation aérienne); the NLR (Nationaal Lucht- en Ruimtevaartlaboratorium); the RLD (Rijksluchtvaartdienst); the LVNL (Luchtverkeersleiding Nederland); the DLR (Deutsches Zentrum für Luft- und Raumfahrt); the DFS (Deutsche Flugsicherung GmbH); the UK CAA (Civil Aviation Authority); the NATS (National Air Traffic Services) and the DERA (Defence Evaluation and Research Agency).

All rights reserved.

## REVISION HISTORY

Date	Revision Number	Reason for revision
Nov 1998		Initial incomplete draft
April 1999	B	Missing parts added
May 1999	1.0	Final version

**DOCUMENT APPROVAL**

	<b>NAME</b>	<b>SIGNATURE</b>	<b>DATE</b>
<b>AUTHOR 1</b>	R. Whitaker		
<b>AUTHOR 2</b>			
<b>AUTHOR 3</b>			
<b>Project Leader</b>	R. Whitaker		
<b>PHARE Programme Manager</b>	H. Schröter		

## EXECUTIVE SUMMARY

The PHARE validation project was formed “to provide a common assessment methodology across all the large scale, real-time simulations conducted by the members of PHARE”. In the execution of this directive, the PHARE validation project:

- Developed a common assessment methodology;
- Produced recommended Measures of Merit for use in evaluating the results from the PHARE Demonstrations;
- Produced recommended analytical methods to be employed in the assessment of the measurements taken.

The experience from the work conducted by the PHARE Validation Project and by the PHARE partners in their application of the methodology, led to the following recommendations:

- The development of a core validation team is central to the successful execution of validation exercises. The team should be involved throughout the complete project lifecycle and its members must have a thorough understanding of the operational concepts under investigation.
- The development of a common Assessment/Validation Methodology is crucial. Traceability of all input and output data is essential to the analysis process. The metrics employed must be clearly defined and relevant to real-life ATM systems.
- The Assessment/Validation Methodology should be seen as “living” and capable of adaptation to the specific requirements of the system under evaluation. Continued development of the methodology is necessary to improve the quality of the output.
- A true validation methodology should not concentrate on one specific technique alone – as within PHARE where real-time simulations only were conducted. The use of other techniques – for example analytical models, fast-time simulators, etc. – is likely to yield more cost-effective results in the early stages of a project’s development cycle.
- For future work similar in nature to the PHARE Programme, a Validation Methodology should be applied as opposed to an Assessment Methodology. Such a methodology should take note of work being conducted within the validation field by both Eurocontrol and the European Commission.

Page intentionally left blank

## LIST OF CONTENTS

<b>EXECUTIVE SUMMARY .....</b>	<b>5</b>
<b>1. INTRODUCTION.....</b>	<b>9</b>
<b>2. COMPOSITION .....</b>	<b>11</b>
<b>3. VALIDATION.....</b>	<b>13</b>
<b>4. COMMON ASSESSMENT METHODOLOGY .....</b>	<b>15</b>
4.1 STAGE 1A – OBJECTIVES .....	16
4.2 STAGE 1B – SELECT ANALYSIS TECHNIQUES .....	16
4.3 STAGE 2 – APPLICATION OF ANALYSIS TECHNIQUES .....	16
4.4 ASSESSMENT AGAINST A BASELINE .....	17
<b>5. MEASUREMENTS .....</b>	<b>19</b>
<b>6. ANALYSIS METHODS .....</b>	<b>21</b>
<b>7. VALIDATION “HISTORY” AND LESSONS LEARNT .....</b>	<b>23</b>
7.1 PD/1.....	23
7.2 PD/2.....	23
7.3 PD/1+, PD/2+, PD/1++.....	24
7.4 PD/3.....	25
7.5 PD/3CT .....	26
7.6 TRAINING .....	26
<b>8. SUPPORT TO PROJECTS .....</b>	<b>27</b>
<b>9. CONCLUSIONS &amp; RECOMMENDATIONS .....</b>	<b>29</b>
<b>10. REFERENCES .....</b>	<b>31</b>

## LIST OF FIGURES

<i>FIGURE 1 : COMMON ASSESSMENT METHODOLOGY .....</i>	<i>15</i>
<i>FIGURE 2 : ASSESSMENT AGAINST A BASELINE .....</i>	<i>17</i>

## LIST OF TABLES

TABLE 1 : DOCUMENTS PRODUCED IN THE VALIDATION PROJECT .....	27
--	----

This page left intentionally blank



## 1. INTRODUCTION

1.1 The tasks of the Validation Project may be summarised as:

**To provide a common assessment methodology across all the large scale, real-time simulations conducted by the members of PHARE.**

1.2 These tasks are further described in the Validation Project Terms of Reference (ToR). These Terms of Reference define the following tasks to be conducted by the Validation Project for all of the large-scale, real-time simulations to be conducted by PHARE members:

### (1) Statement of the Objectives

- (a) To assist in the definition of the goals and objectives of all large-scale, real-time simulations to be conducted by the PHARE members.
- (b) To assist in determining the attributes of the simulated systems to be assessed.

### (2) Advice on Experimental Management

- (a) To assist in the development of an experimental test plan.
- (b) To assist in organising, preparing and conducting all large-scale, real-time simulations to be conducted by the PHARE members

### (3) Provision of Evaluation Methods

To define, select and, where appropriate, develop analytical techniques for:

- (a) Recording and documenting all aspects of the experiment in perusal of the declared aim;
- (b) Analysing the recorded and documented data;
- (c) Presenting the results to PHARE members and the general ATM community.

1.3 The Validation Project was, therefore, responsible for conducting these tasks for the PD/1, PD/2 and PD/3 series of trials, including their respective additional trials – PD/1+, PD/2+, PD/1++. The Validation Project was not responsible for conducting any of those activities in relation to the PD/3 IOCP exercises or to evaluation of the Airborne Project's studies; however, members of the Validation Project were available for consultation by the Airborne Project and IOCP staff on all matters covered by the ToR.

This page left intentionally blank

## 2. COMPOSITION

2.1 The Validation Project has composed the following members from the PHARE Partners during the life of the project:

CENA	R Salvi
	N De Beler
	J Garron*
DLR	F Schick
	H Derkum
	S Tenoort
DRA(Malvern)	C Kelly
	B Bradford
	P Goillau
EEC	A Jackson
	A Drew
	A Marsden*
NATS	R A Whitaker*
	S Kay
	D Hudson
NLR	P Journa
	H Nijhuis
	D van Touw*

\* indicates current Validation members.

This page left intentionally blank

### 3. VALIDATION

- 3.1 The term “validation” has been widely used within the ATM research community over the past few years, each time taking a slightly different meaning. The PHARE Validation Project was initiated before the wider validation debate started, and was concerned more with preparation for and conduct of each trial’s analysis – hence the reference to the provision of a “common assessment methodology” in the task description given in Section 1.1.
- 3.2 During 1998 the EATMS Validation Strategy working group developed a definition of the term validation that is now generally accepted within the ATM community; namely:

**“The process through which a desired level of confidence in the ability of a deliverable to operate in a real-life environment may be demonstrated against a pre-defined level of functionality, operability and performance.”**

- 3.3 As is apparent from the above definition, the role of validation is to build confidence that the ATM system being developed will be fit for purpose. With this in mind, it may be stated that the aims and objectives of the PHARE Validation Project are totally compatible with the above definition; however, the methodology employed is a subset of what may be considered as a complete end-to-end validation methodology.

This page left intentionally blank

#### 4. COMMON ASSESSMENT METHODOLOGY

- 4.1 The development of a common Assessment Methodology has been core to the Validation Projects work. The PD/1 analysis initiated the methodology, and it has been developed and improved throughout the PHARE programme.
- 4.2 The aim of the Assessment Methodology developed by the PHARE Validation Project is to ensure the systematic application of the most appropriate techniques during the PHARE research programmes. It must result in quantitative, impartial and cost-effective assessment of the PHARE solutions to predicted future operational ATC requirements.
- 4.3 The methodology forces the analyst to formulate the objectives of the experiment, and the hypotheses to be tested, prior to starting an evaluation and then to report all substantiated conclusions, together with the scientific methodology employed in the investigation.
- 4.4 The assessment lifecycle applied by the PHARE Validation Project is illustrated below:

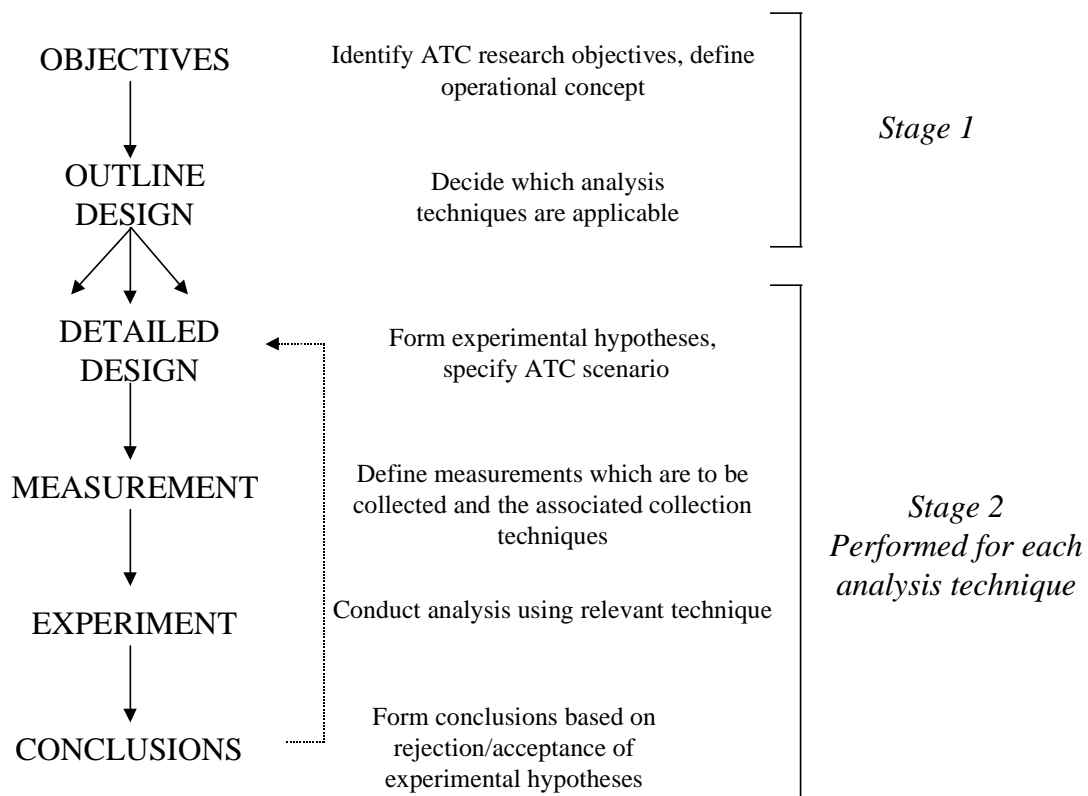


Figure 1 : Common Assessment Methodology

#### 4.1 STAGE 1A – OBJECTIVES

- 4.1.1 The first step in applying the PHARE Validation Assessment Methodology involves: identifying specific research objectives; drawing on the operational requirements; sketching out concepts designed to meet the requirements. This work is, obviously, conducted in co-ordination with other PHARE projects – some elements of which may take full responsibility for specific areas – such as concept development.



#### 4.2 STAGE 1B – SELECT ANALYSIS TECHNIQUES

- 4.2.1 Within PHARE the analysis techniques has concentrated solely on the use of real-time, controller-in-the-loop simulations; with the exception of PD/1++ where fast-time simulations preceded the real-time trial. While the reliance on real-time simulation is not the most cost-effective or necessarily desirable analytical technique, the decision to rely solely on real-time simulations was made in the early days of the PHARE Work Programme and stood throughout.

#### 4.3 STAGE 2 – APPLICATION OF ANALYSIS TECHNIQUES

- 4.3.1 The second stage is to use the analysis technique selected. Experimental hypotheses, for instance that the PHARE operational concept under investigation should reduce the number of conflicts, are developed from the overall research objectives. The details of the ATC scenario to be modelled are defined in terms of experimental parameters, e.g. airspace sectorisation or traffic sample and, for real-time trials, establishing a statistically balanced timetable for the individual experiments.
- 4.3.2 The next step is to decide which measurements are to be made in the experiment, and how to capture these data. The experiment is then conducted, and finally conclusions about the validity of the experimental hypotheses are formed.
- 4.3.3 The final process after the completion of stages 1 and 2 is to collate the conclusions and report the results of the assessment.



#### 4.4 ASSESSMENT AGAINST A BASELINE

4.4.1 Central to the Assessment Methodology used within PHARE Validation is the comparison of results against a baseline system. The studies are specifically designed to assess the difference between a baseline system or mode of operation, and one or more advanced systems or modes of operation. The diagram below illustrates this comparison technique.

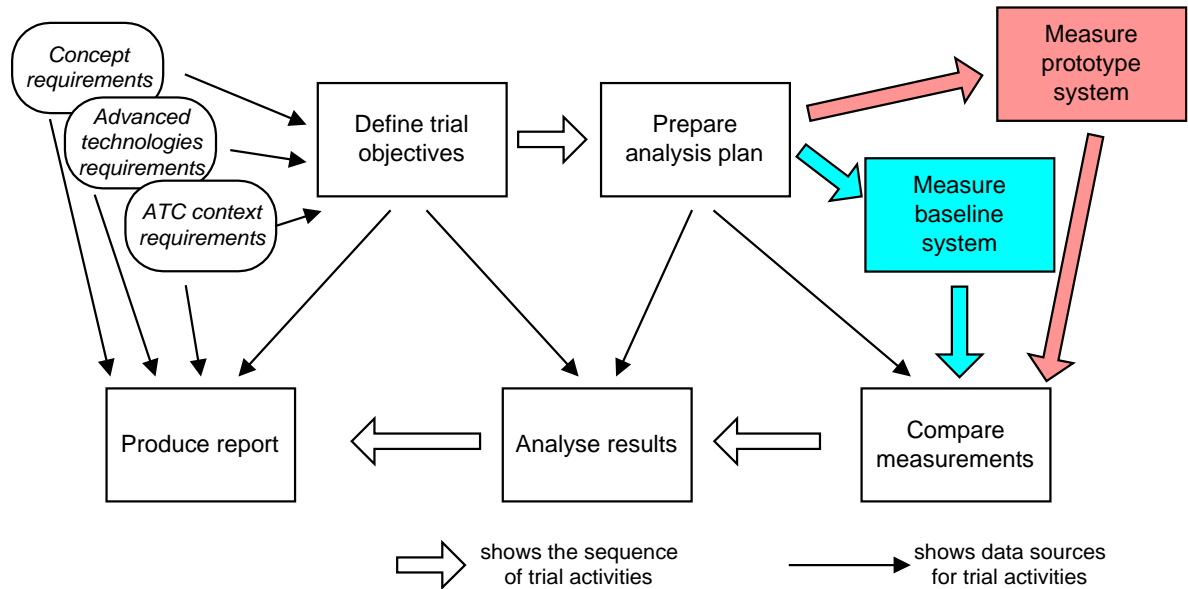


Figure 2 : Assessment Against a Baseline

This page left intentionally blank

## 5. MEASUREMENTS

5.1 Recorded data may be broken down into the following three types, depending on the source of the data:

- Subjective measures include those where the controllers perform some form of self-assessment, questionnaire responses, and comments made by controllers during debrief sessions.
- Objective measures are those which are directly recorded by the simulation system, including video and audio recording of controller activities.
- Observed data is collected by suitably qualified specialist observers during trials. Although subjective in nature, the formal definition of the events that are to be observed and noted by the observers allows much of the subjective interpretation to be eliminated.

5.2 Each of the recorded data types may be collected in a non-intrusive, intrusive or disruptive manner during trials on the PHARE Partners' simulators:

- Non-intrusive data collection is totally transparent to the participating controllers. Data recording through the system software of, for example, aircraft on frequency, is non-intrusive; as is recording of all controller and pilot communications. Non-intrusive data collection will not affect the controllers' performance in any way.
- Intrusive data collection will be apparent to the controllers; however, it should have a minimal affect on the performance of their duties. The collection of workload measures through Instantaneous Self-Assessment (ISA) – where the controller is required to push a numbered key every two minutes (see description below) – is intrusive data collection. Although it may initially distract the controllers from their main duties, they soon become used to performing such a simple task without affecting their overall performance.
- Disruptive data collection, as its name implies, will affect controllers in the performance of their duties by intermittently interrupting the controller. For example, asking the controllers to complete another task in addition to that of controlling the aircraft in their sector whenever there is sufficient time, is disruptive. However, such secondary tasks can be used as an indicator of controller workload. For example, a number might be displayed at regular intervals to the controller, who must then perform a mental mathematical calculation based on that number and input the result via the keyboard. The length of time taken to respond, the accuracy of the result and other such factors are then used as a measure of the controller's workload.

5.3 During PHARE, only non-intrusive and intrusive data collection methods have been employed to date.

5.4 The following sources can be used during the running of trials on the PHARE Partners simulators:

- The logging of objective and subjective measurements associated with controller workload, airspace capacity and quality of service provided to airlines;
- The logging of objective measurements associated with the evaluation of the computer assistance tools and Human Machine Interface (HMI), e.g. the time for which certain tool windows are open on the radar display;
- The use of questionnaires to measure controller acceptance of tools, HMI features and associated technologies such as datalink and operational concepts;
- Observation and video of the actions of controllers during the run by specialist observers;
- Structured debrief sessions after the trial runs, designed to investigate the detailed reasoning behind actions taken by controllers during the run and their opinions on all aspects of the simulation run - including their views on the validity of the measurements taken during the run.

5.5 A full description of the measurements proposed by the PHARE Validation Project for PD/3 may be found in Reference 1.

## 6. ANALYSIS METHODS

- 6.1 The analysis methods applied within the PHARE Validation Assessment Methodology are based on the initial application of suitable statistical analysis. Following a statistical examination of the available data, the analysis process can then take place.

### Statistical Analysis

- 6.2 The type of statistical analysis performed on measurements from PHARE studies falls into two categories:

- Descriptive statistics;
- Inferential statistics.

- 6.3 The use of descriptive statistics is concerned with making concise descriptions of the gathered data. These data may be described by their average, variance, or – say – the proportion of the data above a certain cut-off value. Graphs and histograms may be plotted to show how averages, variances, etc. vary under different experimental conditions. However, the limitation in these methods is that it is not possible to form any conclusions which predict what the results would be in a repetition of the experiment, nor whether any observed differences are genuine changes or, instead, chance effects. The formation of such conclusions requires the methods of statistical inference.

- 6.4 Statistical inference is concerned with two types of problem: estimation of population parameters and tests of hypotheses. Statistical inference allows the experimenter to draw conclusions about a large group of subjects on the basis of measurements from a small sample. For example, in the PHARE Partners' simulators, statistical inference allows a conclusion to be drawn – with a specified level of confidence – that a particular measurement made under the baseline experimental conditions really differs from a measurement made under the advanced experimental conditions. This confidence level relates to the fact that there is always the possibility that a large difference between the measurements simply occurred by chance. Hence, conclusions are stated with an associated probability – i.e. the probability that the observed difference between the measurements of the two systems would have occurred by chance if there was, in reality, no difference between the systems.

- 6.5 Within the PHARE Validation Assessment Methodology, the statement that there is no statistically significant difference between two sets of measurement is stated as a “null hypothesis” (H0). An “alternative hypothesis” (H1) describes a contradiction of the null hypothesis, i.e. that there is a statistically significant difference between the two sets of measurement. The process of statistical inference either accepts the null hypothesis, or rejects it in favour of the alternative hypothesis.

- 6.6 Various statistical tests are available as tools to formalise and standardise the procedures for drawing such conclusions.

### Analysis Methodology

- 6.7 In order to test a number of statistical null hypotheses, the measurements are organised into representative groups. All the data unsuitable for analysis is

then removed. The criteria for exclusion of data must be clearly defined and justified; for example, criteria for unusable data have included:

- Runs which terminated before the intended run duration;
- Runs with equipment problems.

6.8 Such criteria should be set prior to the trial, rather than once the initial analysis results are available.

6.9 The relevant statistical tests can then be performed on the data. MS Excel templates have been developed to perform the most common statistical tests used in the statistical analysis of trial data. Dedicated statistical software packages, such as SPSS and BMDP, are also available to perform additional statistical analysis. The first level in this analysis is simply to compare the descriptive statistics under one experimental organisation to those under another organisation for each traffic sample and controller.

6.10 The second level in the analysis is to conduct tests of statistical inference. The null and alternative hypotheses to be tested are formulated before the experiment in the study design phase. These hypotheses will typically be based on the expected benefit to be shown by the demonstration or on previously proven hypotheses for other associated measurements. On the PHARE Partners' simulators the tests are comparisons of an experimental baseline and advanced system, or of two different advanced systems, rather than an absolute evaluation. The result of the statistical analysis does not allow a conclusion to be reached on the absolute impact of a particular system.

6.11 The selection of the particular statistical tests to be used depends on the exact experimental design used and a careful assessment of the assumptions of the tests available. The selection will also depend to some extent on the size of the experimental dataset from the experiment that is deemed acceptable for statistical analysis.

6.12 A full description of the analytical methods proposed by Validation for PD/3 may be found in Reference 1.

## 7. VALIDATION “HISTORY” AND LESSONS LEARNT

As described in the Introduction, the PHARE Validation Assessment Methodology was built up over the period of the main PHARE Demonstrations, i.e. PD/1, PD/2, PD/3 and their respective additional trials. This section charts the history of the development of the Assessment Methodology based upon these milestones, and examines the lessons learnt.

### 7.1 PD/1

7.1.1 The PD/1 trial was the first conducted and was the responsibility of NATS and its subcontractor, DERA (then the DRA). Although no detailed Assessment Methodology had been developed at this stage a significant amount of preparation was undertaken by the Validation Project. Their work centred on developing the experimental design and on defining the measurements required to be collected for analysis (Reference 2). Controller workload was recognised as a major measure to be collected, and the application of two workload measures was developed for use, namely:

- Instantaneous Self Assessment (ISA);
- NASA Task Load Index (TLX).

7.1.2 PD/1 served as the springboard for detailed development of the Assessment Methodology, and the following significant lessons were learnt;

- Thought must be given to the way in which the recorded data is stored. All the PD/1 data was recorded sequentially on a single file per run, leading to significant problems and wasted time in data retrieval;
- Controller comments must be formally recorded at all times – there were many instances of “so-and-so said such-and-such.....I think”;
- Continuity of personnel involved in the validation process is a must. In PD/1 those conducting the post-trial validation were not necessarily involved in the pre-trial training, nor in the observing or debriefing exercises;
- Those conducting the validation must fully understand the system with respect to its ATC/ATM related operation;
- Those conducting the analysis should also conduct the data extraction (much effort was wasted in the PD/1 analysis due to working on wrong data);
- A full audit trail must be established;
- A well defined analytical methodology must be developed.

### 7.2 PD/2

7.2.1 The PD/2 trial was conducted by DLR who were responsible for the conduct of the validation work. Although no formal document on measurements and

analysis existed prior to the conduct of PD/2, many of the formal methods later adopted for PD/3 were in place within the Validation Project.

7.2.2 Two of the major lessons learnt from PD/1 were taken on board and further developed:

- The need to organise the recording of data into a readily usable format;
- The need to have quick and ready input from a run into the post run briefing.

7.2.3 When addressing the second point, two advances were made over PD/1. During the execution of PD/1, the implementation of the NASA TLX was conducted by a post-trial paper questionnaire. Prior to PD/1+, an electronic version of the NASA TLX had been developed and was implemented during PD/2. At the end of each run, the TLX questionnaire appeared on the controller's radar screen for completion. The controller's inputs were then immediately analysed and the results printed for discussion in the post-run debrief.

7.2.4 A second advancement was in the production of post-run trajectories. Since PD/2 was designed to investigate improvements to the landing schedule at Frankfurt, the efficiency of the aircraft's approach could be visually ascertained from a printout of the trajectories. Again, this was produced immediately following each run and was useful in questioning the controllers as to the procedures that they used.

### 7.3 PD/1+, PD/2+, PD/1++

7.3.1 These three trials were all conducted by NATS. During these trials, the Assessment Methodology was formally applied since it varied little from the normal *modus operandi* of the NATS' Trials Team. Significant lessons that have been learnt during these trials are:

- 1) The application of a well established Assessment Methodology is a pre-requisite for the conduct of any real-time simulation; however, the methodology should be seen as “living” and capable of adaptation to the specific requirements of the system under evaluation;
- 2) Specific documentation should be established as part of the pre-trial validation process, namely:
  - a. Engineering Plan – this details elements such as the layout of the simulator, the ATC tools to be employed, the equipment to be used, etc.
  - b. Analysis Plan – this details elements such as:
    - The trial's aims, objectives, hypotheses.
    - The overall design, including the scenario, traffic samples, the number of exercises to be conducted and their duration, the data to be recorded, the trial organisations to be examined, etc. The trial design is, amongst other things, intended to ensure that the trial is capable of statistical examination.
    - The measurements to be used.
    - The analytical techniques, statistical test to be employed.



- 3) Those responsible for conducting the validation/assessment should be involved throughout the complete project lifecycle;
- 4) The development of well-defined metrics is of extreme importance. The following categories were defined within which metrics were developed:
  - a. Controller Workload
  - b. Quality of Service
  - c. Capacity
  - d. Usability
- 5) While the measurement of capacity was an aim during all the PHARE trials, it has become apparent that no sufficiently robust method exists within the Assessment Methodology for calculating airspace or sector capacity from the data recorded, especially when comparing different airspace designs;

#### **7.4 PD/3**

- 7.4.1 The Assessment Methodology was well defined for PD/3 and formally documented within Reference 1. While the Validation Project was responsible for the definition of the methodology, the individual PD/3 sites were responsible for its execution during and after the trials.
- 7.4.2 Although the application of the full methodology was only applied during the CENA PD/3 trial, the NLR and EEC trials applied those elements of the methodology applicable prior to the trials. All three PD/3 sites commented that the application of the Assessment Methodology helped to guide the development of the planned trials, and was a useful and necessary element of the overall project. In addition to re-learning some of the lessons of the past, the following new points were identified:
  - Predictions should be made of the amount of data to be recorded and the means by which it can be stored and backed-up.
  - The use of intrusive measurements, such as eye-tracking and heart-rate variability, should be considered within the Assessment Methodology.
  - Guidance should be given to the executors of the trial as to how the “ideal” balanced design developed prior to the trial can be modified as the trial is progressed and runs are lost due to – for example – system problems. When some runs are lost decisions must be made as to which runs need to be conducted to ensure that the post-run statistical analysis can be conducted.
  - The role of the observers should be further developed within the Assessment Methodology.
  - Further effort should be applied to measuring controller workload. The primary measures used were ISA and TLX – other areas for workload measurement should be evaluated.

## **7.5 PD/3CT**

- 7.5.1 While the PD/3 Continuation Trial was not technically part of PHARE, it would be expected that the PHARE Assessment Methodology would be applied to this work programme. However, the decision was taken not to conduct any measured trials during PD/3CT; therefore, no discussion can be undertaken on the methodology with respect to this work.

## **7.6 TRAINING**

- 7.6.1 While the training of the controllers should fall within a Validation Methodology, it did not fall within the remit of the PHARE Validation Project in the development of the Assessment Methodology. The effect of controller training on the successful conduct and assessment of the trials was apparent throughout the PHARE Programme. If the controllers are not sufficiently trained and not sufficiently cognisant of the aim of the work, then the results and conclusions drawn are likely to be suspect. As the importance of the training became apparent to the PHARE Project, those responsible for the validation and assessment became more and more involved in the training aspects.

## 8. SUPPORT TO PROJECTS

8.1 The Validation Project has supported all the PHARE trials conducted to date, with the exception of the PD/3 IOCPs. The following has been produced by the Validation Project in support to PHARE:

<b>Title</b>	<b>Reference</b>	<b>Date</b>
PD/3 Measurement and Analysis Specification	PHARE/NATS/VAL-4.4.2/WP005	February 1998
Instantaneous Self Assessment training presentation for use in PD/3	PHARE/NATS/VAL-4.2.6/ISA	August 1997
NASA Task Load Index training presentation for use in PD/3	PHARE/NATS/VAL-4.2.6/TLX	August 1997
Collection software for the NASA Task Load Index		
User guide for the NASA TLX collection software		
Wilcoxon-Mann-Whitney analysis template	PHARE/NATS/VAL-4.2.15/WMW	
Wilcoxon Signed Ranks analysis template	PHARE/NATS/VAL-4.2.15/WSR	
User guide for the Wilcoxon-Mann-Whitney analysis template	PHARE/NATS/VAL-4.2.15/WMW	
User guide for the Wilcoxon Signed Ranks analysis template	PHARE/NATS/VAL-4.2.15/WSR	
Analysis of TLX factors in PD/3	DASR\PHAR-VAL\TN\060	August 1997
Variables for PHARE Demonstrations Evaluation	PHARE Validation Tools Group Working Paper	September 1994
Template of Measurements to be used in PHARE Demonstrations	Eurocontrol DOC 94-70-07	March 1994

Table 1 : Documents produced in the Validation Project

This page left intentionally blank

## 9. CONCLUSIONS & RECOMMENDATIONS

- 9.1 The Validation Project supported all the real-time simulations conducted by the PHARE partners with the exception of the PD/3 IOCPs. During this time an Assessment Methodology has been developed that is compatible with the definition of validation as defined by EVAS. The matured PHARE Assessment Methodology as defined for use within PD/3 is documented in Reference 1.
- 9.2 Through the application of the PHARE Assessment Methodology, the following recommendations can be made for the application of assessment and/or validation to real-time simulations of the type conducted within the PHARE Programme:
- The development of a core validation team is central to the successful execution of validation exercises. The team should be involved throughout the complete project lifecycle and its members must have a thorough understanding of the operational concepts under investigation.
  - The development of a common Assessment/Validation Methodology is crucial. Traceability of all input and output data is essential to the analysis process. The metrics employed must be clearly defined and relevant to real-life ATM systems.
  - The Assessment/Validation Methodology should be seen as “living” and capable of adaptation to the specific requirements of the system under evaluation. Continued development of the methodology is necessary to improve the quality of the output. Specific examples of improvements to the PHARE Assessment Methodology include:
    - Development of a method to examine airspace and sector capacity in PHARE type studies;
    - Examination into the use of intrusive data collection techniques such as eye tracking;
    - Development of the role of the specialist observers.
  - A true validation methodology should not concentrate on one specific technique alone – as within PHARE where real-time simulations only were conducted. The use of other techniques – for example analytical models, fast-time simulators, etc. – is likely to yield more cost-effective results in the early stages of a project’s development cycle.
  - For future work similar in nature to the PHARE Programme, a Validation Methodology should be applied as opposed to an Assessment Methodology. Such a methodology should take note of work being conducted within the validation field by both Eurocontrol and the European Commission.

This page left intentionally blank

**10. REFERENCES**

- 1 PD/3 Measurement and Analysis Specification; PHARE/NATS/VAL-4.4.2/WP005;1.0; February 1998
- 2 Template of Measurements to be used in PHARE Demonstrations; Eurocontrol DOC 94-70-07