

This Document is issued as EATMP Reference Material. The contents are not mandatory. They provide information and explanation or may indicate best practice.

Validation of the Human Error in ATM (HERA-JANUS) Technique

Edition Number	:	1.0
Edition Date	:	12.05.2003
Status	:	Released Issue
Intended for	:	EATMP Stakeholders

DOCUMENT CHARACTERISTICS

TITLE			
Validation of the Human Error in ATM (HERA-JANUS) Technique			
Document Identifier		EATMP Infocentre Reference:	030422-01
HRS/HSP-002-REP-04		Edition Number:	1.0
Abstract			
<p>This report is the third in a series developed within the Human Error in ATM (HERA) Project dealing with how human errors in Air Traffic Management (ATM) can be analysed to improve safety and efficiency in European ATM operations. The purpose of this work is to increase the effectiveness of error recording, analysis and prevention. This report describes the results of two validating exercise using the HERA-JANUS Technique (see EATMP, 2003).</p>			
Keywords			
Human error	Validation	Incident analysis	Inter-rater reliability
Taxonomy	Classification system	Technique	
Contact Person		Tel	Unit
Michiel WOLDRING, Manager, HRS Human Factors Sub-Programme (HSP)		+32 2 7293566	Human Factors & Manpower Unit (DIS/HUM)
Authors			
A. Isaac, M. Lyons, T. Bove and D. Van Damme			

STATUS, AUDIENCE AND ACCESSIBILITY			
Status	Intended for	Accessible via	
Working Draft	<input type="checkbox"/> General Public	<input type="checkbox"/> Intranet	<input type="checkbox"/>
Draft	<input type="checkbox"/> EATMP Stakeholders	<input checked="" type="checkbox"/> Extranet	<input type="checkbox"/>
Proposed Issue	<input type="checkbox"/> Restricted Audience	<input type="checkbox"/> Internet (www.eurocontrol.int)	<input checked="" type="checkbox"/>
Released Issue	<input checked="" type="checkbox"/>	<i>Printed & electronic copies of the document can be obtained from the EATMP Infocentre (see page iii)</i>	

ELECTRONIC SOURCE		
Path:	G:\Deliverables\HUM Deliverable pdf Library\	
Host System	Software	Size
Windows_NT	Microsoft Word 8.0b	

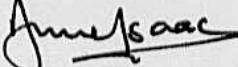
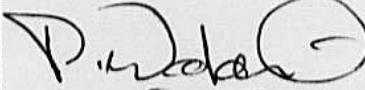
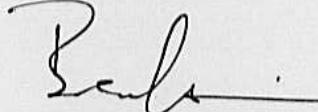
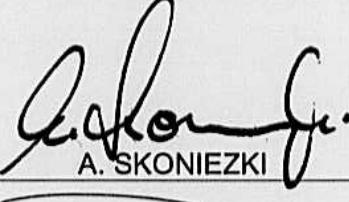
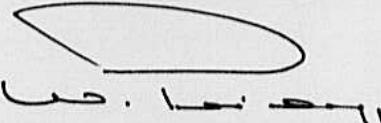
EATMP Infocentre
 EUROCONTROL Headquarters
 96 Rue de la Fusée
 B-1130 BRUSSELS

Tel: +32 (0)2 729 51 51
 Fax: +32 (0)2 729 99 84
 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
<i>Please make sure that the EATMP Infocentre Reference is present on page ii.</i>		
HERA Project Leader	 A. ISAAC	16.05.2003
Chairman HRT Human Factors Sub-Group (HFSG)	 V.S.M. WOLDRING	16.05.03
Manager EATMP Human Resources Programme (HRS-PM)	 M. BARBARINO	16/05/03
Chairman EATMP Human Resources Team (HRT)	 A. SKONIEZKI	16/5/03
Senior Director Principal EATMP Directorate (SDE)	 W. PHILIPP	25.05.03

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	29.05.2000		Working Draft	All
0.2	30.08.2001		Draft	All
0.3	18.02.2002		Proposed Issue for HRT17	All (document configuration)
0.4	31.10.2002		Second Draft for Validation Meeting	All (includes FAA/EUROCONTROL Action Plan 12 results)
0.5	07.02.2003		Second Proposed Issue for HRT19	All (document configuration)
1.0	12.05.2003	030422-01	Released Issue	All (final editorial changes + document configuration)

CONTENTS

DOCUMENT CHARACTERISTICS	ii
DOCUMENT APPROVAL	iii
DOCUMENT CHANGE RECORD	iv
EXECUTIVE SUMMARY	1
1. INTRODUCTION	3
1.1 Overall Work Plan and Focus of this Report.....	3
1.2 Structure of the Report.....	4
1.3 The HERA-JANUS Technique: Taxonomies and Methods of Use.....	4
1.4 Validity	5
1.5 Reliability	6
2. APPLICATION OF VALIDATION TO HUMAN ERROR MODELS	9
3. THE GENERAL METHOD CHOSEN FOR THE PRESENT VALIDATION EXERCISES.....	11
3.1 Introduction	11
3.2 Goals of the HERA-JANUS Validation Exercises	12
4. PART 1 – VALIDATION OF THE HERA-JANUS WORK WITH AIR TRAFFIC CONTROL SUBJECT MATTER EXPERTS	13
4.1 Variables Measured	13
4.2 Exercise Procedures	14
4.3 Results of the Objective Data.....	15
4.4 Results of the Subjective Data	20
4.5 Conclusions	27
5. PART 2 – VALIDATION OF THE HERA-JANUS WORK WITH INCIDENT INVESTIGATORS AND SAFETY MANAGERS	29
5.1 Variables Measured	29
5.2 Exercise Procedures	29
5.3 Results of the Objective Data.....	30
5.4 Results of the Subjective Data	31
5.5 Conclusions	34
6. DISCUSSION OF THE RESULTS	37

REFERENCES	39
FURTHER READING	41
ABBREVIATIONS AND ACRONYMS.....	43
CONTRIBUTORS.....	45
APPENDIX: SPECIFICATIONS FOR THE CASES TO BE PROVIDED FOR THE HERA-JANUS VALIDATION	47

EXECUTIVE SUMMARY

This report is the third in a series developed within the Human Error in ATM (HERA) Project dealing with how the examination of human errors in Air Traffic Management (ATM) can be improved to enhance safety and efficiency in European ATM operations. The purpose of this work is to increase the effectiveness of error recording and analysis. This work has arisen as a result of the increasing recognition of the consequences of human error, error recovery and error reduction in ATM. The effective analysis of incidents in ATM becomes more important as traffic levels increase, as European airspace becomes more harmonised, and as ATM operational centres make more use of computerised support and automation.

This report describes the results of two validation exercises designed to measure the extent to which professional users of the HERA-JANUS Technique (covered by the second HERA Project report - see EATMP, 2003) agree on the errors and associated causal factors as described in authentic ATM incident reports.

The HERA-JANUS validation exercises were carried out in February 2000 and November 2002, and involved 34 professional users. This was one of the most extensive studies ever to be carried out concerning the consistency of error classification in any domain. Therefore, prior estimates of results of the study were very tentative, there being only scant results of a similar nature available.

In addition to describing the objective data from the formal validation exercises, this report describes results from a HERA-JANUS user survey eliciting responses from 43 subjects involved in this work.

The results indicate that there was a high agreement between all subjects in using the technique, independent of whether they were human factors specialists, incident investigators, safety managers or air traffic controllers.

Page intentionally left blank

1. INTRODUCTION

1.1 Overall Work Plan and Focus of this Report

The overall work plan for this project is summarised in Figure 1. This report describes the results of the third of three Work Packages (WPs) which aim to validate a technique for analysing human errors in ATM (see EATMP, 2003).

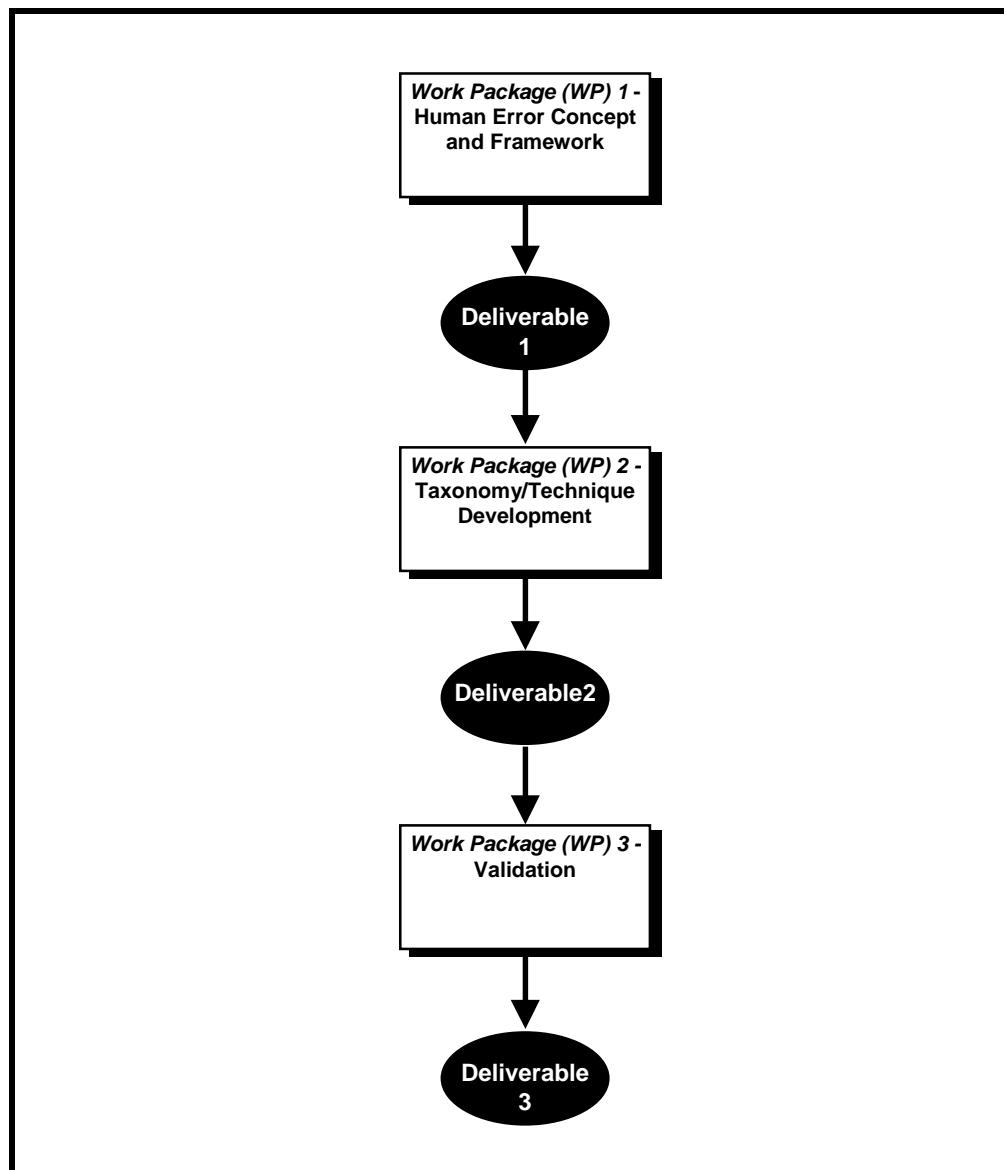


Figure 1: Overall work plan for Phase 1 of the HERA Project (HERA 1)

Work Package 1 successfully managed to identify and develop a human information processing model which was adapted to the present and future jobs in the ATM environment (see EATMP, 2002a, 2002b). This model was then used as the basis for the development of a structured technique (Work Package 2) to allow investigation of ATM incidents (see EATMP, 2003).

Work Package 3, consisting of this report, was designed to analyse how effective this technique would be by way of validation exercises with incident investigators, safety managers, ATC specialists and human factor experts. The report describes these validation exercises, which had the following objectives:

- to determine to what extent the HERA-JANUS Technique was a **reliable instrument** for the analysis and classification of ATM incidents involving human errors as described in incident reports;
- to collect **users' perceptions** of possible shortcomings and inadequacies of the HERA-JANUS Technique and their views of the potential and practical applicability of HERA-JANUS as an analysis and classification tool;
- to **collect data** from the validation subjects' use of the HERA-JANUS Technique to provide any improvements.

1.2

Structure of the Report

The remainder of this report is concerned with describing the goals and the design of the HERA-JANUS validation exercises, the results of the validation exercises, the observations made during the exercises, and finally users' perceptions of the reliability, scope and utility of HERA-JANUS.

The two exercises, held in 2000 and 2002, are reported as Parts 1 and 2 respectively in Sections 4 and 5 of this document.

1.3

The HERA-JANUS Technique: Taxonomies and Methods of Use

In this report the term 'HERA-JANUS Technique' is meant to denote both the **methods** of use involved in applying HERA-JANUS and the **classification** which in turn is composed of several **taxonomic groups**. The HERA-JANUS taxonomic groups are composed of:

- Tasks,
- Equipment,
- Contextual Conditions,
- Error/Violation types,
- Error Detail,
- Error Mechanism,
- Information Processing levels.

The methods of use are outlined in the second HERA deliverable (WP 2) which describes the classification framework (see EATMP, 2003). The validation of the HERA-JANUS Technique is thus a test of both the classification framework, involving more than 350 individual categories organized in seven taxonomic groups, some of which have higher (less detailed) and lower (more detailed) levels, and of the methods of using this comprehensive framework.

1.4

Validity

Validation in this report refers to the testing of a tool, method or technique to ascertain whether it does what it purports to do in a reliable and 'truthful' way. The two main components of the validation exercises were to ascertain whether the technique and method could be used **reliably** and whether the ATM experts could make **valid** decisions about the predication of categories which were in agreement. A robust validation should also be able to provide developers and users with results that indicate the level of reliability and validity that the technique or method may be expected to achieve.

To argue that a technique is **valid** several aspects might be considered. Some of these are described below:

- *A priori* validity - An intuition or common-sense estimate of the content of a test.
- Concurrent validity - Evaluation by comparing the results to known performance evaluated by actual working conditions.
- Congruent validity - Establishing the new method by correlating the outcome with other known valid methods of measuring the same outcome.
- Consensual validity - Assessment based on the number of people who concur about the right outcome.
- Content validity - Systematic examination of the content of the actual items that make up the method by reviewing each item for appropriateness and balanced so that all areas of interest are represented appropriately; considers the relevance of the material included based on expert judgement and is situation specific for circumstances.
- Construct validity - The degree that the method captures the qualities it was intended to capture.
- Convergent/discriminate validity - The degree that the outcome of the method correlates well/poorly with the variables that it should in principle correspond to.

- Empirical validity - The degree to which the method works with real cases in a real sample, often based on comparison to a criterion.
- Face validity - The degree that the method appears to be appropriate. This subjective assessment is typically used only during initial development and resembles *a priori* validation.
- Incremental validity - The added value of the method compared to other methods.

In evaluating a method's validity, one can also discuss the method in terms of its:

- comprehensiveness or how well it captures all characteristics of the overall situation;
- diagnosticity or the degree that the method is able to pinpoint specific sources of error;
- sensitivity, i.e. the responsiveness of the method's output to reflect subtle changes in the input and whether the method responds to minor but potentially important cues;
- Usability, i.e. the convenience and practicality of the method for those who use it and whether they have the capability to use it.

1.5 Reliability

Reliability is often considered 'hand-in-glove' with validity and agreement (consistency) between analysts was also important to the overall goal of the work. To be able to compare data between incidents and to be able to summarise data in trend analyses, it is important that a technique be able to yield similar data when separate incident situations share similar characteristics whether the analysis is done by the same analyst (intra-analyst agreement) or different analysts (inter-analyst agreement). **Intra-analyst agreement**, sometimes called intra-rater reliability, describes statistically the extent to which the same person analysing the same incident (or, in real world terms, a highly similar incident) would come to the same conclusions. **Inter-analyst agreement**, sometimes referred to as inter-rater reliability, describes statistically the extent to which two (or more) people analysing the same incident (or, in real world terms, a highly similar incident) would come to the same conclusions.

Reliability is different from, but related to, validity. For example, it is a generic term to cover all aspects of dependability of a measurement device to yield consistent (same approximate results) when used repeatedly under similar conditions. This can be the degree to which a procedure or instrument is reliable over two testing sessions, resulting in a quantitative expression of

reliability as a reliability coefficient by comparing the two sets of resulting scores.

Inter-rater agreement is a measure of the degree to which multiple coders will classify errors into the same taxonomic categories. A common measure of inter-rater agreement is coefficient Kappa, defined as: where P_o is the proportion of observed agreement among raters, and P_c is the proportion of agreement expected by chance.

$k = (P_o - P_c) / (1 - P_c)$. Kappa ranges from a value of 0 to 1, indicating no agreement and perfect agreement, respectively. Depending on the type of data being coded, acceptable values of kappa typically range from .60 -.74. Values of $k = .40$ or less are considered 'poor' agreement while values of $k = .75$ or greater are considered 'excellent' levels of agreement (Cohen, 1960). Because the categories are assumed to be independent, mutually exclusive, and exhaustive, kappa values can also be interpreted as an indication of the clarity of the category boundaries.

Reports of inter-analyst agreement for other reliability tests of human error identification methods have shown kappa to range from .23 to .95. Reports of intra-analyst agreement have shown kappa to range from .66 to 1.00.

When comparing kappa between studies, careful consideration must be given to the methodology of the studies. The techniques and processes for using this method have differences which may influence the agreement. For example, one study showed how inter-analyst agreement between coders declined as the psychological specificity of the technique increased, thus requiring the analyst to make finer-grained determinations.

Page intentionally left blank

2.

APPLICATION OF VALIDATION TO HUMAN ERROR MODELS

In 1999 EUROCONTROL identified eight requirements for a taxonomy and any technique based on it. These requirements were as follows:

- Firstly, it should be usable by specialists from human factors domains, ATC operators and ATC staff who customarily classify incidents. Users should not be required to have a professional background in human factors or psychology to use the technique.
- Secondly, users should produce high inter-analyst and intra-analyst agreement.
- Thirdly, it should be comprehensive enough to be able to classify all relevant types of ATM human errors and to aggregate them into principle categories.
- Fourth, it should be insightful, that is, able to provide a breakdown of causes and factors (human errors, technical and organisational elements) but must also be able to aggregate similar error forms to determine trends and patterns in the data, leading to more prompt warning of errors, and/or better ways of defending against certain errors.
- Fifth, it should be flexible enough so that future ATM developments would be accommodated.
- Sixth, the database resulting from application of the technique should support a variety of types of queries and analyses.
- Seventh, the taxonomy for the technique should be consistent with approaches in other domains.
- Lastly, application of the technique should provide for the appropriate level of confidentiality and anonymity.

Page intentionally left blank

3. THE GENERAL METHOD CHOSEN FOR THE PRESENT VALIDATION EXERCISES

3.1 Introduction

Therefore, to address the questions listed in Section 2, validation of the HERA-JANUS Method was proposed as a series of activities. The general definition of 'validation' adopted was chosen to be comparable to that used by FAA/EUROCONTROL Research and Development Action Plans. For example, FAA/EUROCONTROL Action Plan 5 defines **validation** as:

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

Realising that the strict definition of validation in the statistical sense was not necessarily suitable for some of the activities planned for the HERA-JANUS project, it was agreed that the process of quantifiable validation data should be adhered to when possible. Therefore, the following general definitions to define the goals for validation were adopted:

- Reliability and Objectivity: Consistency in the HERA-JANUS approach such that two independent investigators would achieve a high degree of agreement in identifying the same causal factors in an incident.
- Content-Related Validity: The ability of the HERA-JANUS approach to capture errors and their causal factors compared to the facilities' existing incident investigation approaches. The HERA-JANUS approach should provide added value beyond the existing processes used by the facilities.
- Empirical Validity: The outputs from the HERA-JANUS approach should relate to operational job performance and potential safety improvements (e.g., training) as viewed by those analysing the incidents and those whose job it is to derive improvement/mitigation strategies, such as safety managers.
- Practicality/Usability: The 'reasonableness' in the use of the HERA-JANUS approach relative to the time required for its use, the amount of effort to analyse and process the incident data, and the level of clarity and understanding in exercising the approach.
- Face Validity and Acceptance: The extent to which incident investigation management, facility investigators, and the controller workforce feel comfortable with the procedures and software application, and the use of the resultant data.

In the following section the specific objectives of the HERA-JANUS validation are described.

3.2

Goals of the HERA-JANUS Validation Exercises

The HERA-JANUS validation exercises were designed with the aim of providing empirical support for the eight requirements listed in Section 2:

1. The first goal of the validation was to determine whether professional subjects or analysts will use the system as intended.
2. The question of reliability across users was one of the two most essential issues of the validation exercises.
3. The question of how much familiarity training users would need to apply the classification reliably and in the way it is intended was an important aspect of the basis for adopting the HERA-JANUS Technique.
4. The question of whether cultural bias would affect the reliability and intended way of classification was also considered. Consideration was given to whether the HERA-JANUS Technique would be applied with an equal degree of consistency and agreement by subjects of different nationalities within the ECAC States, and with incident reports from different countries.
5. The question regarding differences in professional background of the analysts was also investigated. Individuals from incident investigation, human factors, ATM safety, selection and training backgrounds were included in the validation exercises.

Before validation could begin, the technique itself had to be tested and data had to be gathered for the validation activities. To accomplish this, several issues had to be resolved. A sufficient number of people had to be trained to use the technique in a similar way. They then had to use the technique to analyse a sufficient number of cases. Feedback on usability and acceptability had to be solicited from users and safety managers.

4. PART 1 – VALIDATION OF THE HERA-JANUS WORK WITH AIR TRAFFIC CONTROL SUBJECT MATTER EXPERTS

4.1 Variables Measured

The first validation followed a design which would allow analysis of several variables. Subjects were compared on their professional background as well as their familiarity and/or training using the HERA-JANUS Technique. The emphasis on professional background and variations in length of training were made in accordance with two of the explicit goals of the validation, namely to test whether HERA-JANUS would be used in the same way by investigators and other specialists and to estimate the right level of training required.

Subjects in the validation were also recruited from different countries, the purpose of this was primarily to investigate whether the HERA-JANUS Technique could be understood and applied by analysts having different language and cultural backgrounds. Similarly, the incidents used in this validation exercise were randomly selected from two ECAC States (Sweden and UK) and two countries outside Europe (Pan Pacific region) in order to demonstrate the wide applicability of HERA-JANUS.

The classification of individuals in terms of nationality (Swedish, UK and other European¹), professional job function (Incident investigators, human factors specialists, ATM safety, selection and training experts) and their training with the HERA-JANUS Technique (extended or basic) can be found in the table below.

Table 1: Classification of Individuals within the first HERA-JANUS validation

*All numbers given in days

While the output of this validation exercise (dependent variables) consisted in classification results delivered by subjects, the input to the trial (independent variables) were the variations in professional and training

¹ Other European countries were represented via ATM personnel from Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Malta, Portugal, The Netherlands and Switzerland.

factors just described. The classification results identified by the analysts were recorded and the main parameter were defined as the extent to which subjects achieved agreement among themselves on individual classifications of error across all incident occurrences.

4.2 Exercise Procedures

There were three separate but identical validation exercises held at the beginning of 2000. These were held in Sweden (with incident investigators), the UK (with human factor specialists) and Luxembourg (with ATM safety, selection and training experts). Each of the three exercises was divided into a training phase - one day - followed by a practice and testing phase - two days.

There were a total of sixteen cases (from Europe and Australasia) which were used in this validation exercise, giving a total of 34 possible error events to analyse.

Procedure

During the planning of the validation trials it became obvious that no useful data would be collected if subjects highlighted different error events from the incident reports.

From previous work using the HERA-JANUS Technique it was known that, if a disagreement was detected among subjects, it could not be determined whether it was due to different ways of applying HERA-JANUS to the same error items, or the fact that the analysts were focusing on different events or errors within the same incident report. To overcome this problem it was decided that the trial procedure had to ensure that raters were considering the same error items or events so that disagreements could be ascribed to the application of HERA-JANUS and not to the identification of these errors. To ensure this, subjects were given incident reports with the error event already identified. In all cases these pre-defined errors had been identified by incident investigators in the original analyses of the occurrences.

Of the sixteen cases, eight were considered practice cases and the groups were given feedback on their results. The final eight cases which were analysed, were classified as test cases in which no feedback was given to the groups.

Materials

Subjects were provided with a workbook containing copies of the HERA-JANUS taxonomy or classification system ('flowcharts and tables'); the sixteen cases and blank recording forms for each case. It should be noted that the materials which were the basis of this validation exercise were authentic incident reports.

4.3 Results of the Objective Data

4.3.1 Measures of agreement

The results of the validation exercise were measured in the following ways:

1. The percentage agreement per HERA-JANUS taxonomic group for all subjects.
2. The percentage agreement per HERA-JANUS taxonomic group for each expert group (incident investigators, human factors specialists and ATM experts).
3. Results of kappa statistical analysis for determining the significance of the overall agreement with respect to chance.

The values for the percentage agreement figures were calculated as follows:

For each error event and within each HERA-JANUS taxonomic group the most common responses were recorded for all the participants. The number of times these responses were chosen was divided by the number of subjects who responded and this was then converted to a percentage. Once the mean values were calculated for each taxonomic group for each error event, a 'mean of means' was calculated for each group showing the agreement over all the error events.

These calculations were carried out not only within all groups but also between the specialist groups of incident investigators, human factors specialists and ATM experts.

It was then necessary to assess the statistical significance of the agreement and a kappa analysis was carried out on all the data.

For all the results reported it should also be noted that for each of the seven taxonomic groups (as seen in [1.3](#)) there are different numbers of sub-levels which explain human performance characteristics in increasingly specific terms allowing the identification of error causes. For example in the taxonomic group, Contextual Factors, there are eleven high-level groups and over two hundred low-level groups (see EATMP, 2003). Therefore, throughout the results section the taxonomic groups will be referenced to both their high or low level.

4.3.2 Overall agreement per taxonomic group

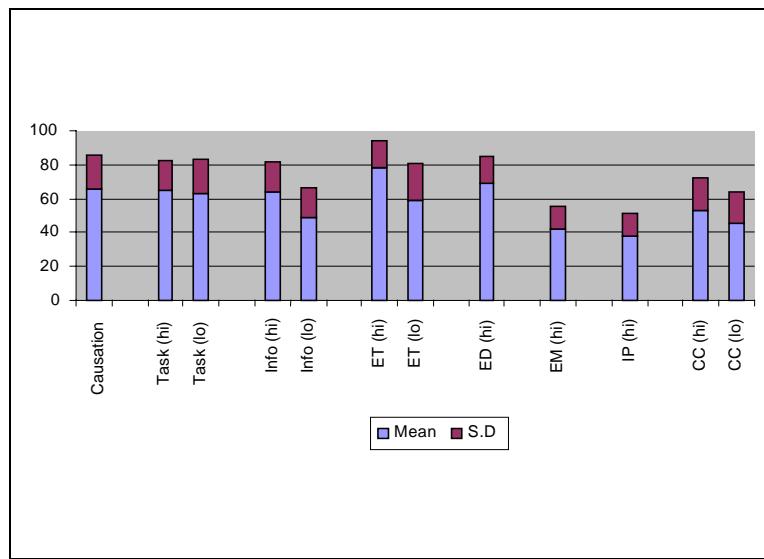
Table 2: Percentage agreement between all subjects

HERA-JANUS taxonomic group	Mean	S.D.
Causation	66	20
Task		
High level	65	18
Low level	63	20
Information		
High level	64	17
Low level	49	18
Error Type (ET)		
High level	78	16
Low level	59	21
Error Detail (ED)		
High level	69	16
Error Mechanism (EM)		
High level	42	14
Information Processing level (IP)		
High level	38	14
Contextual Conditions (CCs)		
High level	53	20
Low level	45	19

*All figures are rounded to the nearest percentage value

Results indicate that the percentage agreement decreases as the level of analysis becomes more detailed within the HERA-JANUS Technique. This happens in the following two ways:

- Firstly, those taxonomic groups which have high and low levels show higher agreement at the higher level. The agreement at the lower level, with a larger number of options available, is less stable. For example, the Information Processing level and the Error Mechanism are both dependent on the choice of Error Detail. Therefore, as expected, the level of agreement obtained in these more detailed levels is lower than that obtained for those less detailed. Incident cases themselves often have few details from the interview if any, which does not help determine a detailed level such as the Information Processing.
- Secondly, when comparing the means it can be seen that the taxonomic groups relating to Task and Contextual Conditions are the groups most affected by the number of levels offered. The Contextual Conditions category has the highest level of detail for analysing an error; therefore, it is not surprising that it has caused the greatest difficulty in identification and agreement with regard to the incidents chosen. Also, these taxonomic groups were the only ones in which more than one choice could be made.

**Figure 2:** Percentage agreement between all participants

4.3.3 Agreement per taxonomic group within the specialist groups

A similar analysis was carried out on the data within the three specialist groups (human factors specialists, incident investigators and ATC experts).

Table 3: Percentage agreement within the specialist groups

HERA-JANUS taxonomic group	Human factors specialists		Incident investigators		ATC experts	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Causation	69	22	73	19	60	17
Task						
High level	70	18	77	20	60	19
Low level	67	24	70	20	68	20
Information						
High level	68	20	70	19	64	13
Low level	60	21	57	21	47	14
Error Type (ET)						
High level	83	17	79	20	78	16
Low level	64	19	64	22	58	20
Error Detail (ED)						
High level	76	19	73	19	63	18
Error Mechanism (EM)						
High level	50	23	51	20	38	16
Information Processing level (IP)						
High level	44	24	46	20	32	13
Contextual Conditions (CCs)						
High level	62	23	60	21	56	16
Low level	57	26	53	23	40	13

*All figures are rounded to the nearest percentage value

From these results it can be seen that the three specialist groups vary in their levels of agreement.

Although in many cases the percentage values are very close, this pattern suggests that with the level of training as given in this study, HERA-JANUS produces more consistency within the groups of human factors personnel and incident investigators than within the ATC experts. In most categories the ATC experts were the least consistent, which may be due to their deeper understanding of the complexity of the tasks and information, and the relevant issues involved. However, it may also be due to the familiarity of the other groups with analytical techniques of this kind.

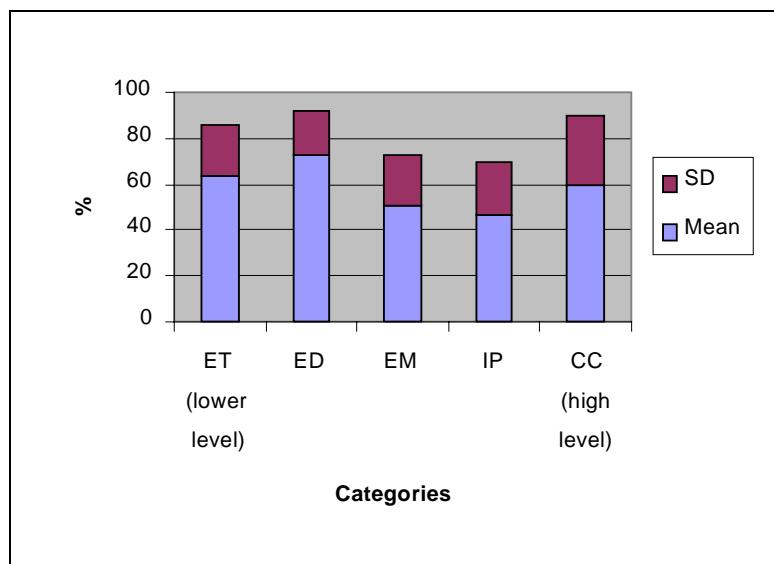


Figure 3: Percentage agreement of the incident investigators

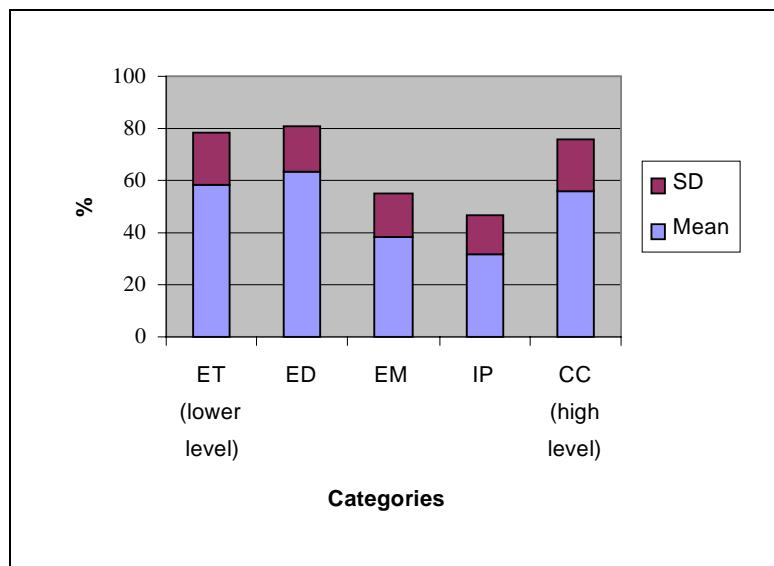


Figure 4: Percentage agreement of the ATC experts

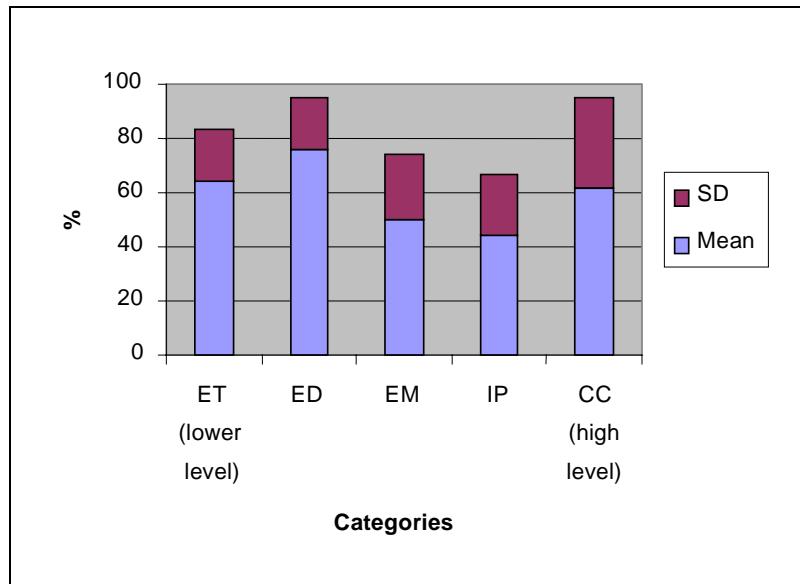


Figure 5: Percentage agreement of the human factors specialists

4.3.4

Results of the statistical analyses

The results shown in the previous sections indicate the agreement between the subjects in percentage terms. However, this does not indicate whether the agreement was significant in statistical terms - that is, the agreement with respect to chance. To identify this the kappa statistic was used. This takes into account the number of options available within the category, the number of analysts and the number of items to be analysed. This produces a value for kappa, which can be converted to a Z value and tested for significance of agreement using the critical values for Z.

The kappa results for each HERA-JANUS category are shown in [Table 4](#).

[Table 4: Kappa Results with regard to the use of the HERA-JANUS Technique](#)

HERA-JANUS taxonomic group	Kappa value	Z value	Significance of agreement*
Error Type (low level)	0.2105	16.87	0.0001
Error Detail	0.3036	32.77	0.0001
Error Mechanism	0.1533	16.42	0.0001
Information Processing Level	0.1322	13.76	0.0001
Contextual Conditions (high level)	0.2106	16.37	0.0001

* Critical Z value for 0.0001 is 3.891

As shown in [Table 4](#), kappa was significant for all the categories analysed. Notably, the highest kappa value was for the Error Detail category and within these choices the lowest one was for the Information Processing level.

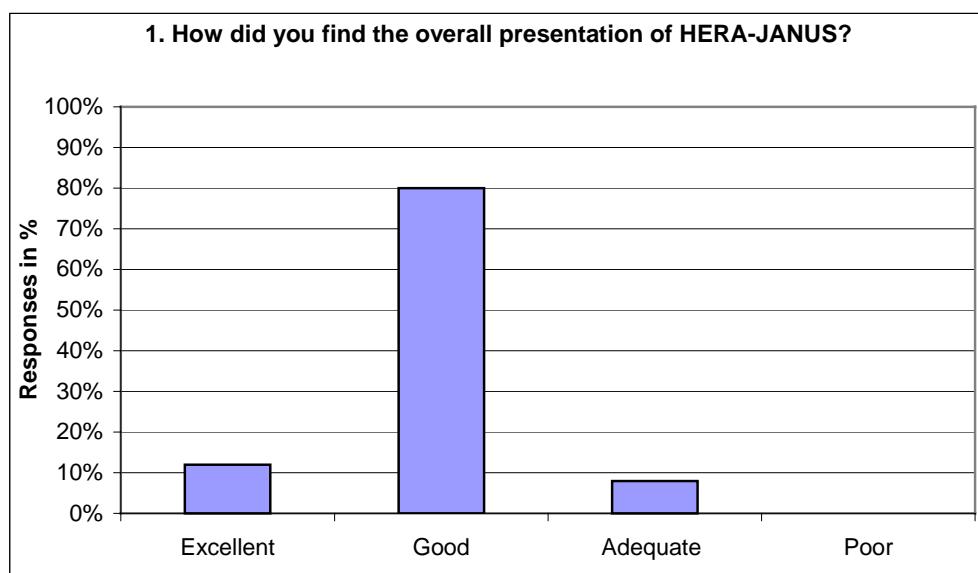
4.4

Results of the Subjective Data

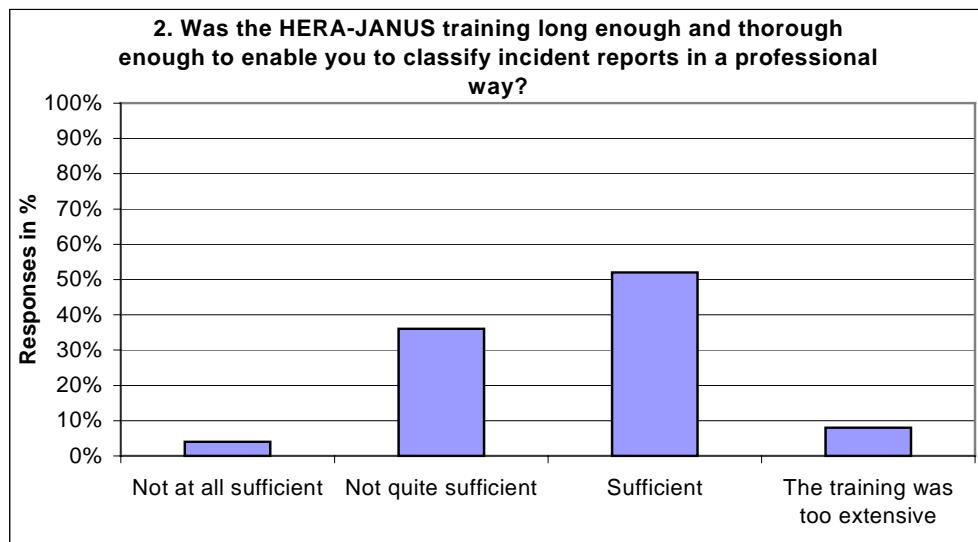
Each of the 27 subjects participating in the validation exercise was asked to give feedback with regard to the validation process. The purpose of the questionnaire was to learn about users' perceptions of a range of issues which concerned the **credibility** and **usability** of HERA-JANUS as well as their views on the validation process itself.

The first two questions were concerned with the written materials and the HERA-JANUS training itself.

Question 1: Overall presentation of HERA-JANUS

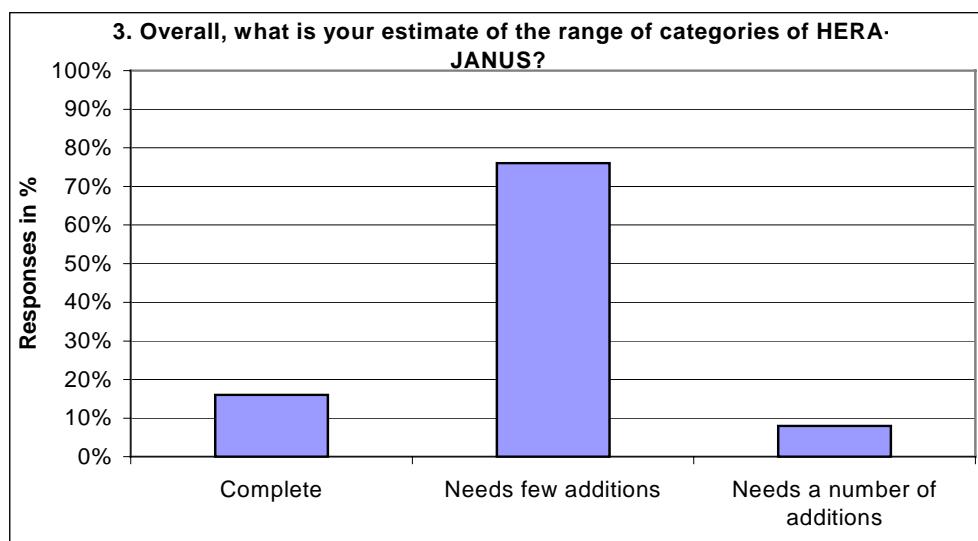


The majority (80%) considered the overall presentation of the HERA-JANUS materials to be good and 11% thought it was excellent.

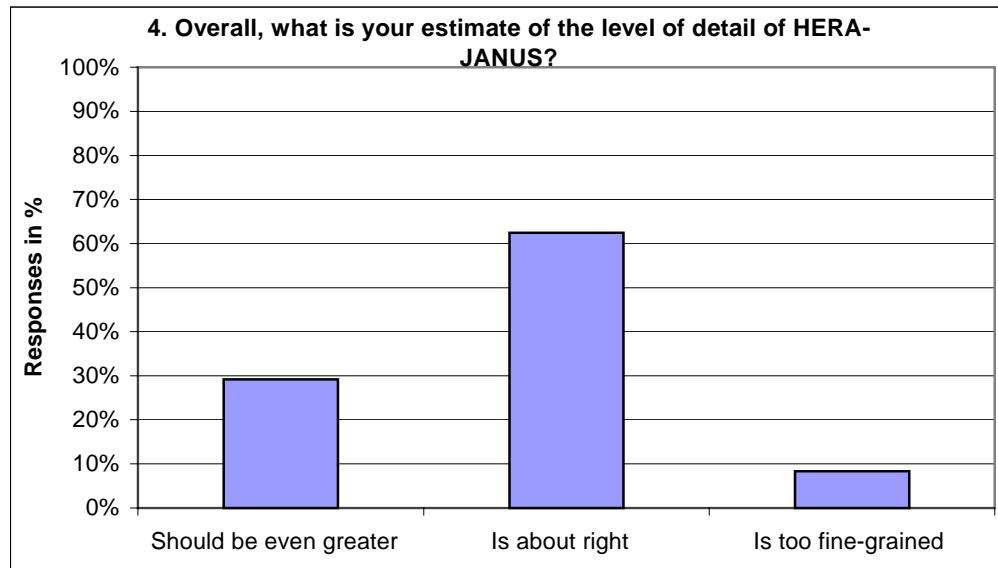
Question 2: Was the HERA-JANUS training sufficient

Although over 50% of the groups considered the training sufficient approximately 40% felt it was not sufficient. When questioned further about this problem, the majority of the groups stated that they believed five days would be sufficient for a novice HERA-JANUS user.

The next two questions were concerned with the range of categories within the taxonomies used in the HERA-JANUS Technique.

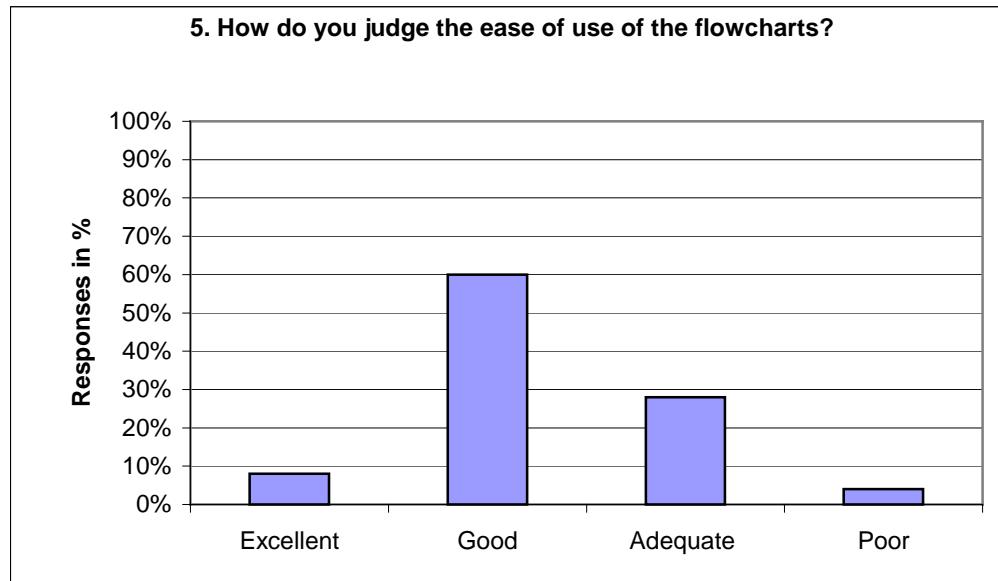
Question 3: What is your estimate of the range of HERA-JANUS categories

The majority of the groups (90%) stated that they thought the categories used were complete or needed few additions. Further discussion suggested that the contextual conditions could be extended and some of the information processing categories could be collapsed. It was also stated that some of the terminologies could be simplified.

Question 4: What is your estimate of the level of detail of HERA-JANUS

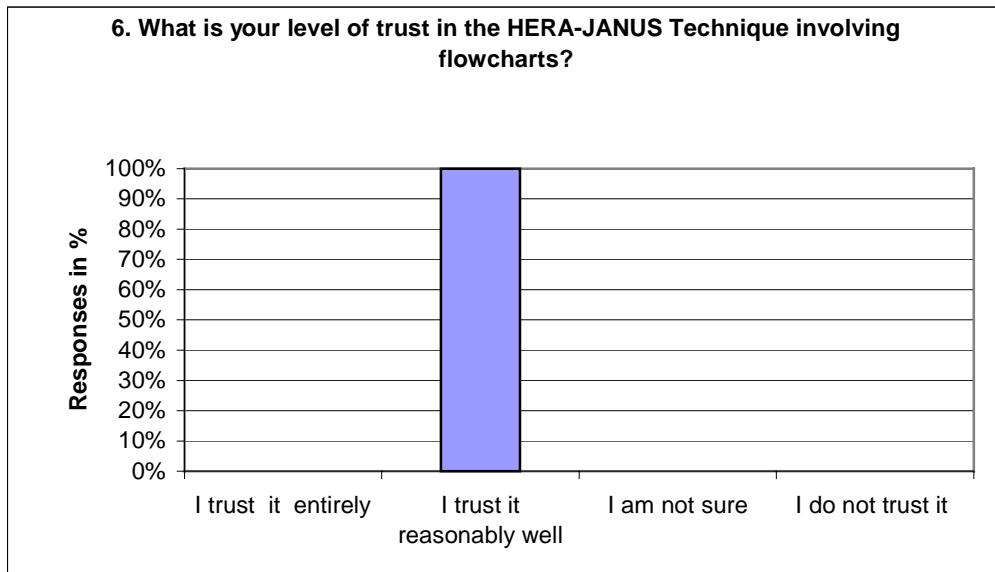
Approximately 60% of the groups considered the HERA-JANUS detail to be adequate.

One of the key items of the questionnaire concerned the validation subjects' impressions of how consistently or reliably they thought HERA-JANUS would be applied in non-experimental situations.

Question 5: Ease of use of the flowcharts

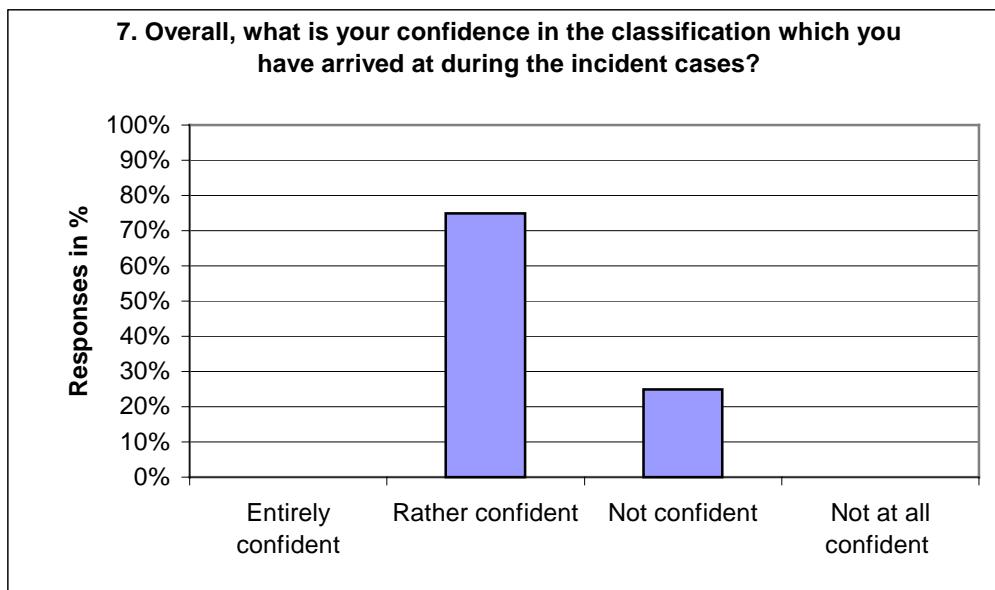
Nearly 70% of the groups thought the flowcharts were good or excellent to use. 29% thought they were adequate and commented that more practice would have been helpful.

Question 6: Level of trust in the HERA-JANUS Technique



All respondents – 100% - commented that they trusted the technique reasonably well.

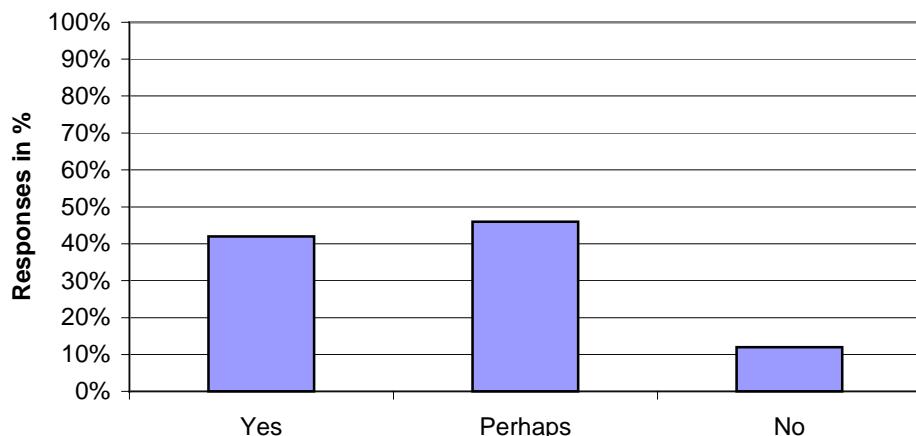
Question 7: Confidence in the classifications of the incident cases



75% of the groups were confident that they had come to the correct conclusions when using this technique. The majority of the respondents who chose the 'not confident' category were those with the least training.

Question 8: Users' perception of HERA-JANUS's reliability

8. If we neglect time constraints involved in training, do you believe HERA-JANUS may be used consistently by different users?

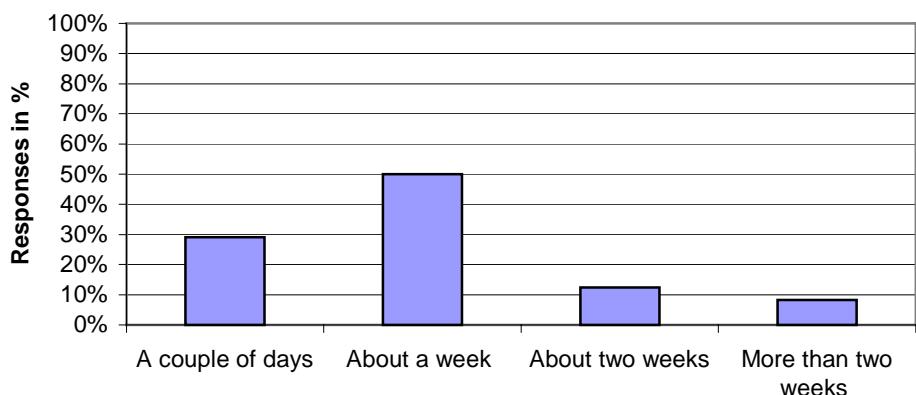


40% of all subjects thought the HERA-JANUS Technique could be used consistently by different users.

A further issue was related to the training requirements for this technique and the issues of understanding of specialist knowledge.

Question 9: The number of days required for training

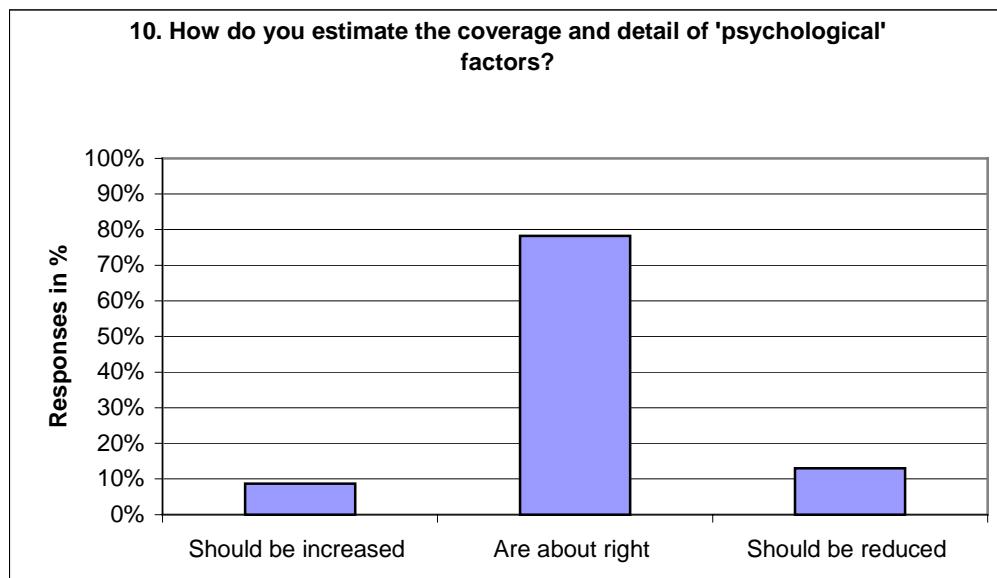
9. On your estimate, how many days would be required to train an ATM professional/investigator to classify incident cases into the HERA-JANUS categories?



Respondents were somewhat divided: 52% stating that one week of training would suffice, 20% estimating two weeks or more, and 28% suggested two days.

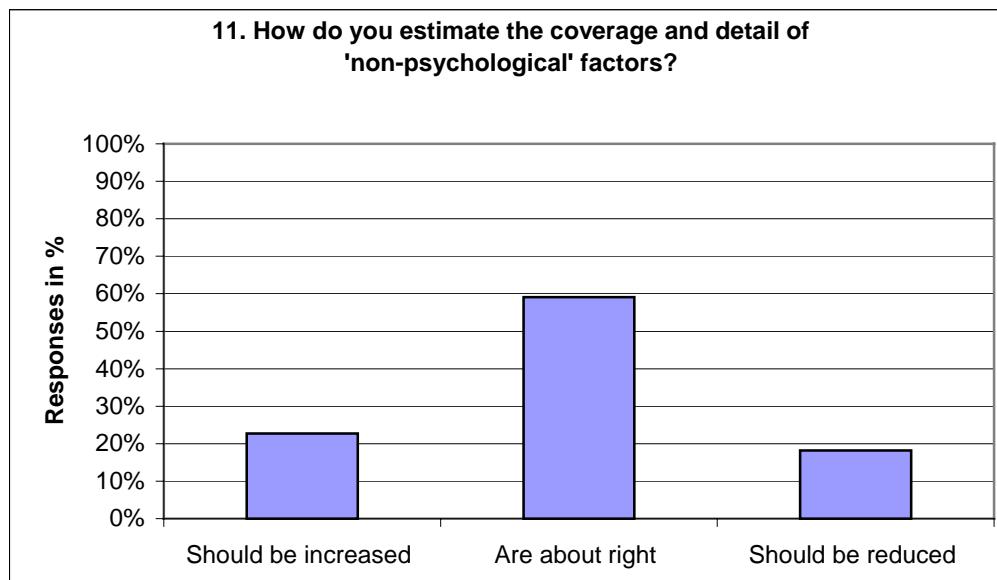
Other questions concerned the level of detail of the HERA-JANUS categories. This item was divided into two separate questions, one concerning the degree of detail of the 'psychological' factors and another the 'non-psychological' factors.

Question 10: Estimate of the coverage of 'psychological' factors



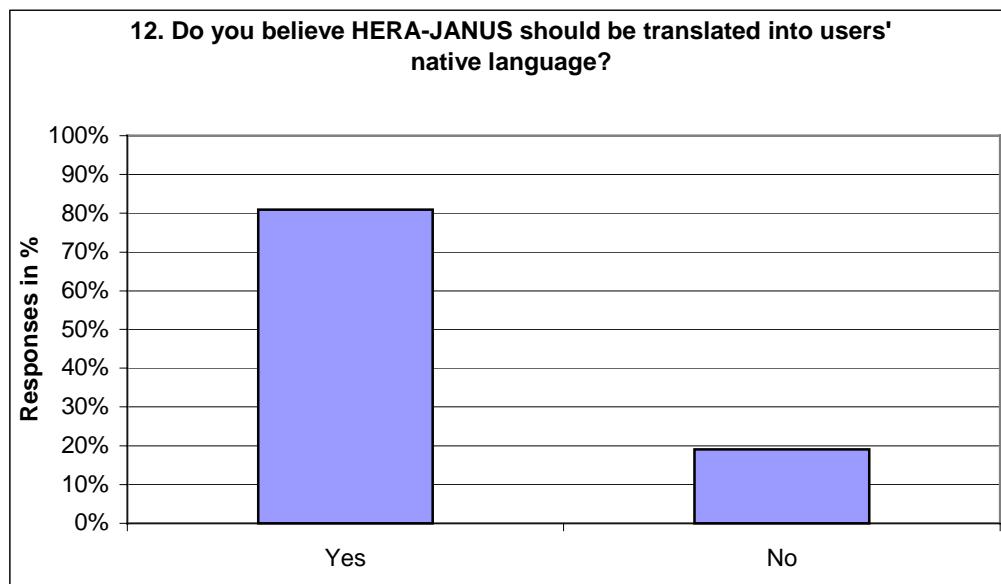
Almost 80% of the groups considered that the detail of psychological factors covered was about right. Other comments, however, indicated that some groups felt that the psychological wording could be simplified.

Question 11: Estimate of the coverage of 'non-psychological' factors



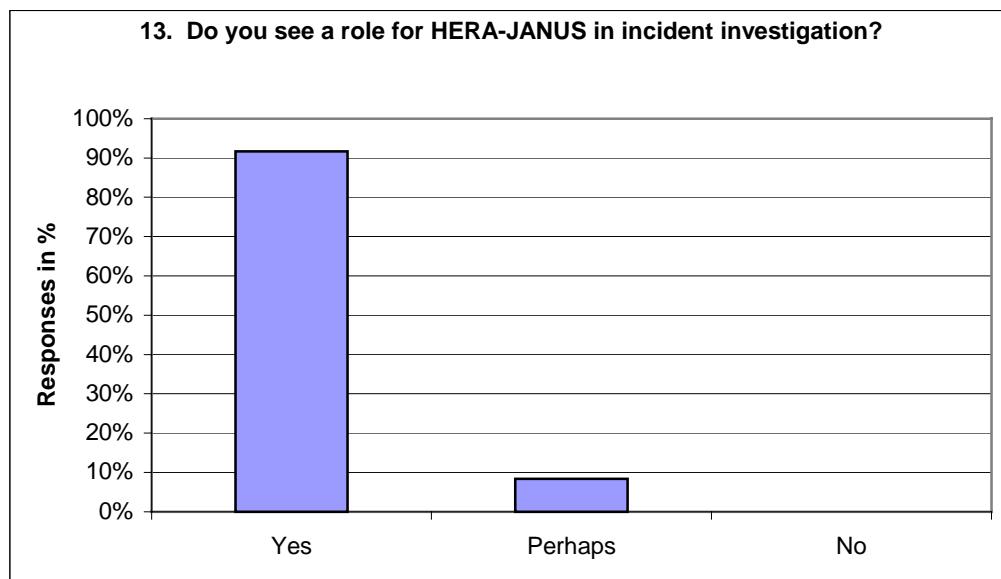
60% of the groups considered that the detail of non-psychological factors covered was about right, whereas 21% felt these details should be increased. Further information with regard to this topic referred to the addition and re-alignment of the Contextual Conditions.

Question 12: The translation of HERA-JANUS into native language



The majority – 81% - indicated that the HERA-JANUS Technique should be translated into the users' native language.

Question 13: The role of HERA-JANUS in incident investigation



Over 90% of the groups felt that there would be a role for the HERA-JANUS Technique in incident investigation.

These responses were particularly interesting in the light of the objective results of the validation. The objective results may be interpreted as indicating that some of the taxonomic groups were too detailed (especially Contextual Conditions). On the other hand, the majority of users reported that they do not seem to wish to reduce the detail of the classification system.

4.5 Conclusions

The original aim of the validation was to find empirical grounds for answering the five questions defined in Section 3.2:

- Firstly, from the high levels of consistency obtained, we can conclude that the users applied the taxonomy as intended.
- Secondly, as verified using statistical analysis, the taxonomy is used fairly consistently among all specialist group analysts. It was also noted that greater consistency was found among those with experience, which is very encouraging. Between specialist groups, it is interesting to note that the highest agreement in each section of the taxonomy was different. The ATC experts had the highest levels of agreement for the Error Type and Error Detail, human factors specialists had the highest levels for Error Mechanism and Contextual Conditions, whereas incident investigators had the highest levels for the Information Processing level. These results only serve to emphasise the importance of incorporating multi-disciplinary groups in incident investigation.
- Thirdly, it is estimated, from both the objective and subjective results, that an optimal training is approximately five days.
- Fourthly, the question of whether the taxonomic group choices were robust across nationalities could not be verified as there were not enough subjects in some national groupings.
- Lastly, it was demonstrated that all groups, including the incident investigators for whom this technique is intended, had a high degree of agreement.

One of the most interesting aspects of the HERA-JANUS Technique that has arisen in this validation work was seen in the information processing level category which indicated the lowest level of agreement between the participants but remained highly significant statistically. This can be explained by the nature of this category. Identification of the information processing categories requires deeper understanding of the cognitive processes influencing the error, and often many incident reports do not record enough detail to facilitate such detailed analysis. However, it may also be that there was relatively more attention paid to the psychological concepts in the analysis, as these aspects are complex for those unfamiliar with them, and as a result little attention may have been directed at the more commonly understood non-psychological elements, such as the Contextual Conditions.

Page intentionally left blank

5. PART 2 – VALIDATION OF THE HERA-JANUS WORK WITH INCIDENT INVESTIGATORS AND SAFETY MANAGERS

5.1 Variables Measured

After approximately fourteen months of ‘beta-testing’ trials and following a ‘beta-testing’ feedback meeting, the second validation exercise was undertaken at the end of October 2002 in the EUROCONTROL Institute of Air Navigation Services in Luxembourg.

Seven participants representing four States attended the validation exercise meeting. The number of participants and the fact that both incident investigators and safety managers were represented allowed for a representative sample to be considered.

5.2 Exercise Procedures

In order to maintain the most robust method possible, only those safety managers and incident investigators who had had full HERA-JANUS training (five days) and who had completed at least seven incident analyses individually were eligible to take part in the validation exercise itself. However, all those subjects who fulfilled the criteria but who could not attend the meeting were sent the subjective questionnaires. Prior to the validation exercise, the safety managers were asked to ensure that at least three original incident cases which had been analysed by the trained investigator, using HERA-JANUS, would be delivered to be used in the validation exercise. A strict protocol of report presentation was given to all participating States (see Appendix). All materials were sent to the exercise coordinator prior to the meeting for copying, and those materials which did not comply with the above format were disregarded.

Having reviewed key academic work associated with inter-rater reliability and expert judgement agreement, three possible candidate statistical analyses emerged. These were Cohen’s Kappa, Kendall’s correlation of concordance and percentage agreement. It was assessed that the first two approaches, which have strict rules of adherence, were unsuitable due to the factors of expertise, experience and homogeneity. Neither of these techniques were therefore suitable and percentage agreement across each subject, case and taxonomy² was therefore used.

² Each taxonomy consists of a variety of alternative options, from groupings of 4 categories to those with a choice of 23 items. Full details of the taxonomies can be found in EATMP (2003)

Procedure

The seven cases used in the validation exercise complied with the format required. The cases represented incidents from four European countries and included a variety of different issues (complexity, functional control area, civil/military and training).

There were seven incident case reports with a total of twenty error events (plus one practice case) presented in random order during the 2 ½ days. After the practice case, delivered by the exercise coordinator, the other incident cases were presented in an identical format. Each State attending the meeting presented at least one incident case for analysis.

Once the factual data of the incident had been presented to the group by the investigators responsible for their analysis, questions were encouraged with regard to the factual issues only. Each investigator was then asked to analyse the incident using the HERA-JANUS Technique entirely on their own. As each investigator completed each case they were encouraged to leave the room and have a break. The investigator responsible for the incident and the coordinator remained in the room at all times. A break of approximately half an hour was taken between each case.

At the completion of all the cases the participants were asked to complete a questionnaire (either in their role as safety manager or incident investigator) relating to all those subjective issues mentioned in the validation questions.

Materials

Participants were provided with the factual data associated with each incident. They also had the HERA-JANUS taxonomies and classifications (flowcharts and tables) and recording forms for their analyses.

5.3

Results of the Objective Data

The average time to present the cases was fifteen minutes and the average time to analyse the cases was 1 hour and 20 minutes.

The total number of errors analysed by the subjects was twenty, with an average of 2.8 errors per incident report (range 2-5).

If any subject did not attempt to complete any section of the analyses, their data was not used for that error event.

The high-level results (given in percentages) can be seen in Table 5.

Table 5: Percentage agreement between case and taxonomic groups

Taxonomy/ Case	Error Type	Error Detail	Error Mechanism	Information Processing level	Contextual Conditions
1	63%	83%	76%	57%	78%
2	100%	100%	58%	92%	65%
3	92%	58%	58%	50%	68%
4	62%	83%	83%	72%	90%
5	83%	75%	75%	58%	83%
6	80%	66%	58%	50%	76%
7	88%	50%	50%	39%	93%
TOTAL	81	74%	65%	60%	79%

The results indicate that despite the complexity of this technique, the incident investigators who were trained and experienced, were able to agree reasonably robustly. The decreasing agreement totals are clearly related to the degree of choice as the taxonomy increases in detail, from the identification of the Error Types to the identification of the Information Processing level involved. One result which indicates some concern in the Information Processing level (case 7) may be explained by the explicitness of this level of analysis. In this case such classifications as the difference between 'failure to integrate information' has to be distinguished from 'failure to consider side effects'. These are complex concepts for human factors specialists and therefore it is not surprising incident investigators have difficulty with these issues. However, the overall percentage agreements per case and taxonomy were very acceptable and lend confidence that the technique is sound and robust.

When analysing the overall agreement of the participants, with each individual incident investigator and their cases, it was indicated that there was 88% agreement.

5.4 Results of the Subjective Data

When analysing the subjective questionnaire responses the following results were ascertained.

Sixteen individuals responded to the request³; four safety managers and twelve incident investigators. Ten of the incident investigators work at national level and the remainder at local level. The average number of years of specialist investigation/safety experience was three and a half years, and

³ This included all those participants at the validation exercise meeting, and other incident investigators and safety managers who were trained and experienced, but who could not attend the meeting.

eight of the subjects had formal training for their position. The average time the incident investigators had been using the HERA-JANUS Technique was between one and one and a half years. The group had analysed an average of seven cases.

The participants were asked about the positive and negative aspects of the HERA-JANUS Technique.

There were 31 positive statements about the HERA-JANUS Technique, 18 negative statements and two participants who did not answer these questions. The statements were clustered in the following groups:

Positive aspects of the HERA-JANUS Technique	Negative aspects of the HERA-JANUS Technique
<ul style="list-style-type: none"> ▪ The technique offers a logical, structured and complete framework, and a more objective approach to the analysis of all components of the incident situation. – 10 responses. ▪ The technique will lead to a better harmonisation of the incident investigation process allowing comparison, data and experience sharing between units and countries, as well as trend analyses and statistical data processing – 3 responses. ▪ The technique is a great support in preparing and conducting the interview with all the persons involved – 10 responses. ▪ The technique allows the inclusion and in-depth analyses of the human factor issues within the investigation process – 6 responses ▪ The technique is the only existing tool to analyse human errors – 1 response. ▪ The technique could contribute to encouraging a blame-free culture within the organisation – 1 response. 	<ul style="list-style-type: none"> ▪ It takes too long to be trained and get familiar with the technique – 4 responses. ▪ It is difficult to identify/define the error – 4 responses. ▪ The terms to be used must be clearly defined – 2 responses. ▪ The use of the technique takes a lot of time – 2 responses. ▪ The technique is not usable for all cases, for instance in which technical problems are more important – 2 responses. ▪ The recommendations from the analysis should be refined – 1 response. ▪ It is difficult to use as a non-native English speaker – 1 response. ▪ It is difficult to understand the model – 1 response. ▪ One needs a human factors background – 1 response.

The results demonstrated that there was more common agreement in the positive aspects, that is more agreement regarding fewer issues, and less

agreement associated with the negative aspects, that is more issues with only one or two respondents for each.

The participants were equally divided in regard to the length of training required for the HERA-JANUS Technique, but most responded that between three and ten days would be appropriate, depending on prior experience.

Once trained 89% reported that they felt comfortable using the technique having analysed between four-ten cases, the average being seven cases.

Most of the participants reported that they had used the HERA-JANUS Technique both in investigation teams and individually.

When asked about the comparison between the HERA-JANUS Technique and their previous incident investigation methods the individuals responded in the following way:

Question	Response	Comments
1. Which method takes more time to use?	<ul style="list-style-type: none"> ▪ HERA-JANUS: 86% ▪ No answer: 12% 	<ul style="list-style-type: none"> ⇒ Half of the group mentioned this was only during training
2. Which method gives you more quantitative information?	<ul style="list-style-type: none"> ▪ HERA-JANUS: 100% 	<ul style="list-style-type: none"> ⇒ High human factors value ⇒ Takes you to the real problem ⇒ Gives more detail
3. Which method gives you more qualitative information?	<ul style="list-style-type: none"> ▪ HERA-JANUS: 86% ▪ Same as other method: 14% 	<ul style="list-style-type: none"> ⇒ You don't miss valuable information ⇒ Gives better interview ⇒ Gives more points of view ⇒ It makes you look for the information
4. Which method helped you better to gather the information?	<ul style="list-style-type: none"> ▪ HERA-JANUS: 86% ▪ No answer: 14% 	
5. Which method helped you identify the human error better?	<ul style="list-style-type: none"> ▪ HERA-JANUS: 100% 	<ul style="list-style-type: none"> ⇒ It directs you to the human errors ⇒ It is very precise on HF issues ⇒ The flowcharts assist very well

Question	Response	Comments
6. Which method was better at supporting the interview process?	<ul style="list-style-type: none"> ▪ HERA-JANUS: 100% 	<ul style="list-style-type: none"> ⇒ It is such a structured approach ⇒ It helps with better questions ⇒ It involves the controller more
7. Which method was better accepted by the operational staff involved in the incident?	<ul style="list-style-type: none"> ▪ HERA-JANUS: 50% ▪ Same as other method: 25% ▪ No answer: 25% 	<ul style="list-style-type: none"> ⇒ It helps to get the correct findings ⇒ There is better involvement of controllers ⇒ Some questions are too personal
8. Which method helped you better with the recommendations?	<ul style="list-style-type: none"> ▪ HERA-JANUS: 75% ▪ Same as other method: 12% ▪ Other methods: 12% 	<ul style="list-style-type: none"> ⇒ It provided additional recommendations ⇒ With more human factors in the report there are less recommendations needed ⇒ Involves management more ⇒ Provides more precise recommendations
9. Which method gives you more confidence in the investigation process?	<ul style="list-style-type: none"> ▪ HERA-JANUS: 66% ▪ Other methods: 11% ▪ No answer: 11% ▪ Don't know: 11% 	<ul style="list-style-type: none"> ⇒ It was more explicit ⇒ It helped explain the human error ⇒ It helps the safety culture ⇒ It is more structured ⇒ It is more logical ⇒ It is more acceptable for controllers

5.5

Conclusions

86% said the HERA-JANUS Technique gave better qualitative results by being more detailed, objective, structured and precise. 100% stated that it gave better quantitative results because it generated more useful human factors information in the interview process and prompted investigators to look in more detail at the context in which the errors had been made.

86% reported that this technique helped to collect incident data and all the subjects agreed that the technique supported the identification of the errors in an incident.

66% reported that it had given them more confidence in the investigation process, particularly the interview activities and that nearly 50% had commented the controllers involved in the investigation of their incidents accepted the HERA-JANUS methodology better than previous methods.

All the subjects stated that they would recommend the use of the technique and stated such things as: "The technique takes an intensive look behind the incident and helps to eliminate the possible causes from the probable facts" or "It replaces the feeling of guessing with a structured approach".

However, six members of the beta-testing and validation exercises mentioned the following improvements:

- the development of a computerised version;
- more guidelines regarding the introduction of the technique in existing investigation processes;
- more support for the recommendations after the analysis;
- more support to use the technique in serious incidents;
- translation into local languages;
- the development on a more concise version;
- the development of a CBT package.

Page intentionally left blank

6. DISCUSSION OF THE RESULTS

It is clear from the validation work that this technique has been one of the most thoroughly tested tools for the investigation of human errors within incidents in any high-reliability system.

The HERA-JANUS system is a complex but comprehensive technique; therefore the HERA-JANUS validation was also very comprehensive, spanning more than 200,000 potential data points. It is therefore not surprising that results of the validation are not simple - different parts of the HERA-JANUS Technique yield different results when used by different groups.

The development of the technique has followed two major steps:

- Firstly, the original incident analysis technique, which was subject to validation with several ATM professional groups. This led to several changes, mainly in the area of the terminologies, and the presentation and simplification of some of the flowcharts. At this time the joint work with the development of the technique was also being undertaken with the FAA in the USA. There was therefore a further demand to test and validate the technique more rigorously with the target audience, that is incident investigators.
- This led to the second validation exercise with trained and experienced incident investigators and safety managers.

The results of the two validation exercises have revealed that the HERA-JANUS Technique has fulfilled most of its goals to produce a useful, valid and reliable method of human factors incident investigation. It seems to be a comprehensive, robust and well-liked technique which has been adopted by several European countries as a valuable tool in safety management.

The question about the extent to which HERA-JANUS may be used across different ECAC States may be usefully addressed in a continued development of HERA-JANUS; in particular, it would be valuable to know how HERA-JANUS will be used in other parts of Europe. Users' reactions to the item about translation of HERA-JANUS are mainly in favour of national versions, and it might be informative to learn how a translation to, and adoption by, different cultures will occur.

Page intentionally left blank

REFERENCES

EATMP Human Resources Team (2002a). *Technical Review of Human Performance Models and Taxonomies of Human Error in ATM (HERA)*. HRS/HSP-002-REP-01. Edition 1.0. Released Issue. Brussels: EUROCONTROL.

EATMP Human Resources Team (2002b). *Short Report on Human Performance Models and Taxonomies of Human Error in ATM (HERA)*. HRS/HSP-002-REP-02. Edition 1.0. Released Issue. Brussels: EUROCONTROL.

EATMP Human Resources Team (2003). *The Human Error in ATM Technique (HERA-JANUS)*. HRS/HSP-002-REP-03. Edition 1.0. Released Issue. Brussels: EUROCONTROL.

Page intentionally left blank

FURTHER READING

Bove, T. (2002). *A human error management taxonomy in air traffic control*. Unpublished doctoral dissertation, University of Roskilde.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Education and Psychological Measurement*, 20(1).

Cook, T.D. & Campbell, D.T. (1979). *Quasi Experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin Co. Boston.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York: Wiley.

Hollnagel, E. (1993a). *Human Reliability Analysis: Context and Control*. London: Academic Press.

Hollnagel, E. (1993b). The phenotype of erroneous actions. *International Journal of Man-Machine Studies*, 39, 1-32.

Isaac, A. & Pounds, J. (2002). *The development of a Trans-Atlantic approach to the analysis of Human Error in ATM*. Paper given at the ATM 2001 Conference, Santa Fe, USA.

Kirwan, B. (1992). Human error identification in human reliability assessment. Part 2: Detailed comparison of techniques. *Applied Ergonomics*, 23 (6), 371-381.

Kirwan, B. (1994). *A Guide to Practical Human Reliability Assessment*. London: Taylor & Francis.

Pounds, J. & Isaac, A. (2002). *Development of an FAA-EUROCONTROL Technique for the Analysis of Human Error in ATM*. Washington, DC: FAA Office of Aviation Medicine Technical Report DOT/FAA/AM-02/12.

Rasmussen, J. (1982). Human errors: a taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents*, 4, 311-335.

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. New York: North Holland.

Reber (1985). *Dictionary of psychology*. London: Penguin.

Shorrock, S.T. (1997). *The Development and Evaluation of TRACER: A Technique for the Retrospective Analysis of Cognitive Errors in Air Traffic Control*. MSc (Eng) Thesis: The University of Birmingham, September 1997.

Page intentionally left blank

ABBREVIATIONS AND ACRONYMS

For the purposes of this document, the following abbreviations and acronyms shall apply:

A/C	Aircraft
ACAS	Airborne Collision Avoidance System
AIS	Aeronautical Information Services
ATC	Air Traffic Control
ATCO	Air Traffic Controller / Air Traffic Control Officer (US/UK)
ATM	Air Traffic Management
ATS	Air Traffic Services
CAA	Civil Aviation Authority/Administration
CCs	Contextual Conditions
CENA	Centre d'Etudes de la Navigation Aérienne (France)
DFS	Deutsche Flugsicherung GmbH (Germany)
DGAC	Direction Générale de l'Aviation Civile (France)
DIS	Director(ate) Infrastructure, ATC Systems & Support (EUROCONTROL Headquarters, SDE)
DIS/HUM	See 'HUM (Unit)'
EATCHIP	European Air Traffic Control Harmonisation and Integration Programme (now EATMP)
EATMP	European Air Traffic Management Programme (formerly EATCHIP)
ED	Error Detail
EM	Error Mechanism
ENAV	Ente Nazionale di Assistenza al Volo (Italy)
ET	Error Type

FAA	Federal Aviation Administration (<i>USA</i>)
FPS	Flight Progress Strip
HERA (Project)	Human Error in ATM (Project) (<i>EATMP, HUM, HRS, HSP</i>)
HRS	Human Resources Programme (<i>EATMP, HUM</i>)
HRT	Human Resources Team (<i>EATCHIP/EATMP, HUM</i>)
HSP	Human Factors Sub-Programme (<i>EATMP, HUM, HRS</i>)
HUM	Human Resources (Domain) (<i>EATCHIP/EATMP</i>)
HUM (Unit)	Human Factors and Manpower Unit (<i>EUROCONTROL Headquarters, SDE, DIS; also known as DIS/HUM</i>)
IANS	Institute of Air Navigation Services (<i>EUROCONTROL, Luxembourg</i>)
IP	Information Processing level
LVNL	Luchtverkeersleiding Nederland (ATC <i>The Netherlands</i>)
NATS	National Air Traffic Services Ltd (<i>UK</i>)
OJTI	On-the-Job-Training Instructor
REP	Report (<i>EATCHIP/EATMP</i>)
R/T	Radiotelephone or radiotelephony
TCAS	Threat and Collision Avoidance System
TRACER	Technique for Retrospective Analysis of Cognitive Errors in ATC
WP	Work Package

CONTRIBUTORS

<u>NAME</u>	<u>ORGANISATION, STATE</u>
-------------	----------------------------

REVIEW GROUP

Dr. S. BAKER	CAA UK
Ms. S. FIGAROL	CENA, France
Prof. E. HOLLNAGEL	Linköping University, Sweden
Mr. B. RUITENBERG	LVNL, The Netherlands

AIR TRAFFIC CONTROL SUBJECT MATTER EXPERTS

HUMAN FACTORS TEAM	NATS, UK
INCIDENT INVESTIGATION TEAM	CAA Sweden
Ms. A. I. CAMBRIAIA	NAV Portugal
Mr. B. CONSIDINE	EUROCONTROL IANS
Mr. J. DELAURE	BELGOCONTROL, Belgium
Mr. R. DOORN	LVNL, The Netherlands
Ms. M. HINTZ	EUROCONTROL IANS
Mr. G. LE GALLO (<i>until 2001</i>)	EUROCONTROL Headquarters
Mr. I. MANSOLAS	EUROCONTROL IANS
Mr. P. NORBJERG	ATC, Denmark
Mr. P. PANKNIN	DFS, Germany
Ms. J. POUNDS	FAA, USA
Ms. J. ROTHAUG (<i>until June 2002</i>)	EUROCONTROL Headquarters
Mr. C. RIZZO	EUROCONTROL IANS
Mr. M. SCHOLTZE	ENAV, Italy
Mr. D. SHEEN (<i>until November 2000</i>)	EUROCONTROL Headquarters

INCIDENT INVESTIGATORS AND SAFETY MANAGERS

Mr. W. ARTNER	AUSTROCONTROL, Austria
Mr. H. AHRENS	DFS, Germany
Mr. I. BALKE	DFS, Germany
Mr. V. BRIDE	DGAC, France
Mr. S. CHAPELIER	DGAC, France
Mr. P. ENGELEN	LVNL, The Netherlands
Mr. G. FALB	AUSTROCONTROL, Austria
Mr. H. HOUTMAN	LVNL, The Netherlands
Ms. K. JONAS	AUSTROCONTROL, Austria
Mr. R. JORGENSEN	CAA Sweden
Ms. P. KELLETER	DFS, Germany

INCIDENT INVESTIGATORS AND SAFETY MANAGERS (*continued*)

Mr. J.-P. KERLEROUX	DGAC, France
Mr. R. KNOSPE	DFS, Germany
Mrs. I. LAPORTE-WULLENS	CENA, France
Mr. J. LEONHARDT	DFS, Germany
Mr. J.-Y. LE LUDUEC	DGAC, France
Ms. A. PECHHACKER (<i>until 2002</i>)	AUSTROCONTROL, Austria
Mr. M. POLEMAN	LVNL, The Netherlands
Mr. K. REISCHL	AUSTROCONTROL, Austria
Mr. B. ROYNE	CAA Sweden
Mr. A. SIDDOWAY	SERCO, UK

Document Configuration

Mrs. C. HELLINCKX (<i>External contractor</i>)	EUROCONTROL Headquarters
---	--------------------------

APPENDIX: SPECIFICATIONS FOR THE CASES TO BE PROVIDED FOR THE HERA-JANUS VALIDATION

1. Framework / Context

Within the EUROCONTROL/FAA Research Programme, Action Plan 12, several European Member States have agreed to contribute to the 'HERA-JANUS' Harmonised Technique beta-testing and validation phases.

Main purposes of the beta-testing phase are:

- to test the usability and acceptability of the HERA-JANUS Technique,
- to gain some lessons from its application and use,
- to collect proposals for its refinement/amendments.

The validation process will be based on new incident cases analysed during the beta-testing phase and that will be provided to the exercise coordinators by the contributing States.

In order to ensure standardisation of the data provided for each case, the following specifications have been defined:

- All information should be provided in English.
- Specifications should be seen as a common framework for the content and the set of data needed.
- Because each case is very specific, not all detailed data may be relevant.
- It should be kept in mind that as much relevant data as possible should be provided, knowing that those involved in the validation will need to understand and identify what happened, how and why it happened, and what were the contextual conditions.
- A simple way to know what is relevant for the incident under consideration would be to look at the contextual conditions that have been identified with HERA-JANUS. For example, if the complexity of the airspace has been identified as a contextual condition within your HERA-JANUS analysis, it would be interesting to describe in few words what makes the airspace sector complex.
- In the writing/re-creating of the incident scenario you should avoid the use of wordings such as "the controller failed to identify the A/C because he had no decision or plan", in order not to bias the validation. The purpose of the validation is to ensure that, with the same information, different analysts will come to the same conclusions.

- It is believed that the report should be approximately fifteen pages long. For the quality of the validation process it would be better to have shorter but well-documented cases.

The participants to the validation phase will be those from European States (incident investigators and safety managers) who have been part of the beta-testing phase and are available to attend the validation meeting.

Participants will need to evaluate how much detail can be kept in their cases in order to ensure compliance with their confidentiality policy.

All data can be dis-identified, for example:

- A/C can be numbered (A/C1, A/C2, etc.), or named by their type (the B747, the DC10, etc.) or by a modified callsign (ABC11, XWZ22);
- the persons involved can also be numbered (P1, P2, P3) or named by their function (flow controller, approach controller, radar controller, planner controller, flight data assistant, trainee, On-the-Job Instructor (OJTI), supervisor, etc.).

EUROCONTROL will ensure that:

- data provided will remain confidential;
- the use of the data will be restricted to the purpose of this scientific research;
- all reports will be destroyed after the validation activity.

2. Case Data

If possible, and relevant to the understanding of the occurrence, the following data should be included:

1. Introduction
2. Time Line of the Incident
3. Factual Data
 - 3.1 Airspace/airport/traffic
 - 3.2 Personnel involved in the incident
 - 3.2.1 Air traffic controllers
 - 3.2.2 Other ATM personnel
 - 3.2.3 Pilots
 - 3.3 Aircraft and flight data
 - 3.4 Meteorology
 - 3.5 Documentation and Procedures
 - 3.5.1 Documentation
 - 3.5.2 Procedures
 - 3.6 Equipment / Control room/tower / Workstation/HMI
 - 3.6.1 Equipment

- 3.6.2 Control room/tower
- 3.6.3 Workstation/HMI
- 3.7 Environment
- 3.8 Communication
 - 3.8.1 Air-ground communication
 - 3.8.2 Ground-ground communication
- 3.9 Organisational aspects
 - 3.9.1 Teamwork
 - 3.9.2 Rostering - shift work
 - 3.9.3 Staffing - training
 - 3.9.4 Unit/team culture
- 4. Conclusions

Annex 1: Radar Plots
Annex 2: Communication Transcripts
Annex 3: Other Relevant Data
Annex 4: Recommendations
Annex 5: HERA-JANUS Analysis Form

For each item, data from different sources - R/T communication, interviews, incident recreation, radar replay, survey, etc. - can be compiled.

3. Case Description

1. Introduction

The introduction should provide:

- a short description of the data collection process (how, with whom, etc.);
- a simplified sketch (drawing) of the sector and adjacent sectors highlighting routes, traffic flows and any important feature such as usual entry/exit points, flight levels, etc.;
- a summary of what happened;
- the A/C involved;
- the person(s) involved;

2. Time Line of the Incident

The detailed time line of the incident should include all useful information for the understanding of what happen.

3. Factual Data

In this part, as much relevant factual information as possible should be provided on the following subjects.

3.1 *Airspace/airport/traffic*

Some general information about the airspace and airport should be provided, such as:

- complexity of the airspace,
- type of traffic,
- restricted airspace,
- airspace structure,
- routes, etc.

Some specific information about the traffic conditions just before and at the time of the occurrence should be provided, such as:

- traffic in the sector at the time, and fifteen minutes before the incident;
- traffic load (percentage of maximum capacity);
- the complexity inherent to the situation, etc.

3.2 *Personnel involved in the incident*

3.2.1 Air traffic controllers

Relevant information should be provided, such as:

- age,
- qualification (licence/rating),
- experience on job/position (in years),
- working hours during the last week (present roster, overtime, etc.),
- time on position before the incident,
- physical fitness/well-being,
- personal information.

3.2.2 Other ATM personnel

The same information as in Section 3.2.1 should also be provided for the following personnel, including any information concerning actions or non-actions that contributed to the incident:

- supervisor,
- OJTI,
- other ATS personnel,
- airport personnel,
- technicians,
- AIS personnel.

3.2.3 Pilots

As factual information/data about the pilots (age, experience, etc.) may be difficult to collect, any interesting information that can be extracted from the R/T communication (language problem, distraction, technical problem, etc.) or the flight information, or any other relevant information should be provided, including any pilot action or non-action that contributed to the incident.

3.3 *Aircraft and flight data*

General information concerning the following should be provided:

- A/C:
 - type,
 - category,
 - equipment (ACAS, TCAS);
- The flight plan.

Any relevant information about specific event/fact at the time of the incident should be provided, i.e.:

- A/C technical problem,
- flight history (re-routing, delay, etc.),
- pilot request, etc.

3.4 *Meteorology*

Any information about meteorological conditions, before or at the time of the incident, should be provided.

3.5 *Documentation and Procedures*

3.5.1 Documentation

General information on the documentation given to the controllers/personnel related to the incident, should be provided.

Specific information about any aspect of the documentation that was involved in the incident, should be provided, i.e.:

- unavailable or out-of-date documentation,
- new documentation,
- relevant information from documents/files,
- unclear information.

3.5.2 Procedures

General information on usual and unusual procedures related to the incident should be provided.

Specific information on any aspect of the procedures that was involved in the occurrence of the incident should be provided, i.e.:

- procedure inappropriate to the situation,
- new procedure recently implemented,
- unclear procedure,
- procedure which is routinely not applied.

3.6 *Equipment / Control room/tower / Workstation/HMI*

3.6.1 Equipment

General information on technical aspects or characteristics on the controller's working position equipment should be provided, i.e.:

- radar screen,
- support tools,
- FPS devices,
- auxiliary equipment.

Specific information of any status of the equipment that contributed to the incident should be provided, i.e.:

- radar failure,
- newly implemented system.

3.6.2 Control room/tower

General information on the control room / tower environment should be provided, i.e.:

- layout,
- noise,
- sources of distraction.

3.6.3 Workstation/HMI

3.7 *Environment*

3.8 *Communication*

3.8.1 Air-ground communication

The R/T communication transcripts relating to the incident should be provided in annex.

Any additional information that could be derived from the R/T communication should be provided, i.e.:

- number of A/C on the frequency,
- use of phraseology,
- technical problem/failure,
- language problem,
- tone of the voice,
- assertiveness of the controller or pilot,
- reactivity of the pilot.

3.8.2 Ground-ground communication

If available, the telephone call transcripts (or a summary of main communication contents translated into English) should be provided.

Any information about other verbal and non-verbal communication between the persons involved should be provided.

3.9 *Organisational aspects*

3.9.1 Teamwork

General information about teamwork or working organisation relevant to the incident should be provided, i.e.:

- team composition (radar, planner, flight data assistant);
- task sharing between team members;
- coordination within the team and with adjacent sectors/units;
- team spirit / atmosphere within the team, and between the team and other teams.

Specific information on any aspect of teamwork that was involved in the incident should be provided, i.e. training or examination issues.

3.9.2 Rostering - shift work

General information about shift work and the rostering system which might be related to the incident should be provided, i.e.:

- shiftwork cycles,
- rostering system,
- working hours, time on position, break organisation,
- rules concerning overtime limits.

Specific information of any aspect of working hours that was involved in the incident should be provided, i.e.:

- the controller requested to change his initial shift role for personal reasons;
- the controller was on position for three hours because ...

3.9.3 Staffing - training

Any staffing information such as temporary or permanent ATCO shortage and causal factors should be provided, e.g. temporary staff shortage due to a specific project requesting ATCO's contribution, and training duties.

3.9.4 Unit/team culture

Any information concerning attitude towards safety, other teams or other Units should be provided, e.g. any known cultural or corporate differences between two Units/Centres.

4. Conclusions

A short conclusion can be provided.