

# VALIDATION OF THE JANUS TECHNIQUE: CAUSAL FACTORS OF HUMAN ERROR IN OPERATIONAL INCIDENTS

*Julia Pounds, FAA, Oklahoma City, Oklahoma, USA*

*Anne Isaac, Eurocontrol, Brussels, Belgium*

## Introduction

This paper reports on work jointly conducted by Eurocontrol and the Federal Aviation Administration (FAA) as part of Action Plan 12 called the Management and Reduction of Human Error in Air Traffic Management (ATM). Only a high level description of the project is presented here. More detailed results will be discussed during the conference presentation.

Human error has been identified as a dominant risk factor in safety-oriented industries such as air traffic control (ATC). However, little is known about the factors leading to human errors in current ATM systems, in particular those human errors contributing to violations of separation standards.

The first step toward prevention of human error is to develop an understanding of when and where it occurs in existing systems and the system variables that contribute to its occurrence. Once these human and system variables are better understood, appropriate interventions can be more specifically defined. This understanding depends on the availability of informative and diagnostic data spanning from the individual to system levels. For example, meaningful data about individual behaviour can be used to manage programs designed to enhance individual performance, such as skills training, decision aiding, and human-centred automation. Likewise, data about factors that influence performance, such as sector characteristics, traffic flow, operational procedures, and teamwork can be used to better manage these elements to mitigate their impacts on individual and thus system performance.

To develop this type of data, two existing approaches to human error identification techniques – the Human Error Reduction in ATM technique (HERA) [1] and the Human Factors Analysis and Classification System (HFACS) [2] – were harmonised. This work resulted in an integrated technique called JANUS. The harmonisation work is described in Isaac and Pounds [3] and Pounds and Isaac [4]. Strengths of the JANUS technique include use of a structured interview process so that psychological errors contributing to the air traffic controller's behaviour can be identified and lessons learned from the incident.

Originally conceived as a method to retrospectively analyse existing incident reports, the technique also showed potential as an investigation tool, having encompassed several categories relevant to human error investigation: Error Detail (ED)—the cognitive domain of the error, e.g., perception; Error Mechanism (EM)—the cognitive function that failed, e.g., detection of information; Information Processing (IP)—the psychological process, e.g., tunneling; and Error Type (ET)—how the error was manifested, e.g., a required action was omitted. These behaviours are viewed as occurring in a dynamic situation that unfold in a sequential and temporal manner rather than looking only at behaviour at the moment separation is lost. Contextual Conditions (CC) which shape performance, such as weather conditions, airspace characteristics, traffic load, and pilot actions are also captured. Further, the event is viewed within its operational environment, including characteristics associated with teamwork, supervision, and the overall organization.

Although these categories had been demonstrated in other studies separately to be important to understanding causal factors related to human performance [1, 2], the

harmonized technique underwent beta testing by seven European nations and the FAA. The results of this beta test were used to validate the technique. The processes of beta testing and validation are described in this report.

## **Background**

The goal of this phase of Action Plan 12 was to test whether the technique would facilitate the extraction of data that is meaningful to aviation safety systems. The purpose of the test and validation of JANUS was to provide an empirical basis that confirms subjective opinion and assesses its added value in relation to previous investigation methods used.

Any useful human error framework should be valid and reliable for the domain of interest, that is, broadly applicable and comprehensively reflecting an accurate picture of human errors in ATM.

### ***Validity***

Assessing the validity of a technique can take several forms [5], such as a priori, concurrent, congruent, consensual, content, construct, convergent/discriminate, empirical, face, and incremental validity. In evaluating a method's validity, one can also discuss the method in terms of its:

- Comprehensiveness, or how well it captures the full range of characteristics in the situation.
- Diagnosticity, or the degree that the method is able to pinpoint specific sources of error.
- Sensitivity, or 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, or the convenience and practicality of the method for those who use it and whether they have the means to use it.

### ***Reliability***

Reliability is often considered hand-in-glove with validity. The reliability of a method is determined by the consistency with

which it can be used—the extent that its use yields same the approximate results when used repeatedly under similar conditions. Consequently, agreement (consistency) between analysts was also important to the overall goals of the project. To compare data between incidents and to summarise data in trend analyses, it is important that a technique yields similar data when separate incident situations share similar characteristics whether the analysis is done by the same analyst (intra-analyst agreement) or by 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. That is, inter-rater agreement is a measure of the degree to which multiple coders will classify an error into the same taxonomic categories.

Several measures of agreement exist, so when selecting and comparing measures of agreement between studies, careful consideration must be given to the goals and methodologies of each [6]. That is, the measures and processes for using them have important differences that may influence differences in agreement. For example, one study showed how inter-analyst agreement between coders declined as the psychological specificity of the classifications increased, thus requiring the analyst to make finer-grained determinations [7]. A common measure of inter-rater agreement is coefficient Kappa, defined as the proportion of observed agreement among raters related to the proportion of agreement expected by chance [8, 9]. Equally useful measures include correlations of concordance, odds ratios, and raw agreement indices, among others [6].

Both validity and reliability are necessary. Neither alone is sufficient. It was

possible that the technique might meet multiple validity criteria but not be used reliably. Users of the harmonized technique should be able to use the tool similarly to extract relevant information and the information should be consistent over similar situations.

## **Application of Validation to Human Error Models**

Kirwan, 1992 [10] identified several potential criteria to be considered in relation to validating human error models. These can also be applied to validating the JANUS technique. The technique should have applicable theoretical underpinnings, be comprehensive, facilitate analyst agreement, show high usability, expedite resource usage, be based on a clear and repeatable procedure, and be acceptable to users of the technique.

Eurocontrol [1] identified eight requirements for a taxonomy and any technique based on it. These requirements were as follows. First, it should be usable by specialists from human factors domains, as well as ATC operators and AT 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. Third, 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 permit the aggregation of 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” (p. 26). 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.

## **General Validation Method**

This study was developed to answer several basic questions related to validity: 1) Does JANUS work? 2) How well does JANUS work? 3) Is JANUS better than the current method? 4) Is JANUS ready for implementation? 5) Will the results from JANUS help to improve safety management?

To address these questions, validation of the JANUS method was proposed as a series of harmonized activities. The general definition of ‘validation’ adopted was chosen to be comparable to that used by other FAA/Eurocontrol 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 a real-life environment may be demonstrated against a pre-defined level of functionality, operability and performance.”

Realizing that the strict definition of validation in the statistical sense was not necessarily suitable for some of the activities planned for the JANUS project, it was agreed that the process of quantifiable validation of the 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 JANUS technique 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 JANUS method to capture errors and their causal factors compared to the facilities’ existing incident investigation approaches. The JANUS technique should provide added value beyond the existing processes used by the facilities
- Empirical Validity: The outputs from the 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 JANUS process 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.

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 consistently apply the technique. 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. An approach to identify additional requirements (refinements and supplementary tools) also needed to be included in the process.

Validation activities posed unique challenges to both the FAA and Eurocontrol. For example, organizational differences in labor-management relationships impacted each study differently. Based on discussions of the differences, it was decided that parallel and complementary approaches be used based on the particular requirements of each to conduct the validation activities. Eurocontrol invited interested member states to volunteer to join this phase of the JANUS development. This included the briefing of the safety managers, the training of the incident investigators and the use of the technique within the every-day investigation process of the member states.

In contrast, the FAA adopted a ‘go-team’ approach. That is, researchers trained in the technique responded to actual events and collected data, which was then used during the validation phase.

## **European Validation Exercise**

After approximately 14 months of ‘beta-testing’ trials and following a ‘beta-testing’ feedback meeting, the validation exercise was undertaken at the end of October 2002 at the Institute of Air Navigation Services in Luxembourg.

### ***Participants***

Seven representatives from four States participated in 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 of involved personnel.

### ***Protocols***

To maintain the most robust method possible, only those Safety Managers and Incident Investigators who had participated in full HERA-JANUS training (5 days) and who had individually completed at least 7 incident analyses were eligible to take part in the validation exercise. However, all those individuals who fulfilled the criteria but who could not attend the meeting were sent the JANUS technique assessment questionnaires.

Prior to the validation exercise the Safety Managers were asked to ensure that at least three original incident cases, which had been analysed, using HERA-JANUS, by the trained investigator would be delivered to the validation exercise co-ordinator<sup>1</sup>. A strict protocol of report presentation was given to all participating States. All materials were sent to the exercise co-ordinator prior to the meeting for duplication. Materials that did not comply with the above format were disregarded.

### ***Method***

There were seven incident case reports (plus one practice case) presented in random order during the 2 ½ days. After the practice case was delivered by the exercise co-ordinator, the other incident cases were presented. Each State attending the meeting presented at least one incident case for analysis.

---

<sup>1</sup> The second author

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 a case they were encouraged to leave the room and take a break. The investigator responsible for the incident and the co-ordinator remained in the room at all times. A break of approximately ½ 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 the validation questions.

### Results

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

The average time to present a case was 15 minutes and the average time to analyse a cases was 1 hour and 20 minutes.

The total number of errors analysed by the participants was 20, with an average of 2.8 errors per incident report (range 2-5).

If any participant did not attempt to complete a section of the analyses, their data were not used for that error analysis.

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 determined that the first two approaches, which have strict rules of adherence, were unsuitable due to the factors of expertise, experience and homogeneity. Percentage agreement across each participant,

case, and taxonomy<sup>2</sup> was therefore used. The high level results (given in percentages) can be seen in Table 1.

**Table 1. Results of the European Validation. Percentage Agreement Across Participants, Case, and Taxonomy.**

<i>Tax.</i>	<i>ET</i>	<i>ED</i>	<i>EM</i>	<i>IP</i>	<i>CC</i>
<i>Case</i>					
<i>1</i>	63	83	76	57	78
<i>2</i>	100	100	58	92	65
<i>3</i>	92	58	58	50	68
<i>4</i>	62	83	83	72	90
<i>5</i>	83	75	75	58	83
<i>6</i>	80	66	58	50	76
<i>7</i>	88	50	50	39	93

Key: *ET* – Error Type, *ED* – Error Detail, *EM* – Error Mechanism, *IP* - Information Processing level, *CC* – Contextual Conditions

The results indicate that despite the complexity of this technique, the incident investigators who were trained and experienced were able to reach reasonable levels of agreement. The decreasing agreement is 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. The only category which indicates some concern is in the Information Processing level where 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 experts and therefore it is not surprising that incident investigators have difficulty with these issues.

<sup>2</sup> 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) (1).

However, the overall percentage agreements per case and taxonomy appear promising.

Fifteen individuals responded to the request to complete the JANUS technique assessment questionnaires: three safety managers and 12 incident investigators. Nine of the incident investigators worked at a national level and the remainder at the local level. The average number of years of specialist investigation/safety experience was three and a half years, and eight of the participants had formal training for their position. When analysing the subjective questionnaire responses the following results were ascertained.

When asked about the comparison between the HERA-JANUS technique and their previous incident investigation methods, eighty-five percent said the HERA-JANUS technique gave better qualitative results by being more detailed, objective, structured and precise. Seventy five percent stated that it gave better qualitative results because it generated more useful information in the interview process and prompted investigators to look in greater detail at the context in which the errors had been made.

Eighty five percent reported that this technique helped to collect incident data. All the participants agreed that the technique supported the identification of the errors in an incident.

Eighty three percent reported that it had given them more confidence in the investigation process, particularly the interview activities. Nearly seventy percent commented that the controllers involved in the investigation of their incidents accepted the HERA-JANUS methodology better than previous methods.

Each of the participants 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” and “it replaces the feeling of guessing with a structured approach”.

## **FAA Validation Exercise**

The FAA exercise was separate from and complemented the European activity. It relied on data developed from interviews with operational personnel after an OE was recorded. The data collection activity is first described, followed by the validation activities. The data collection ran for nine months and twenty-nine air traffic control facilities volunteered to participate. Data were collected in parallel with but separate from the existing FAA investigation process. Facility personnel coordinated the interviews, which were then conducted by researchers traveling to the data sites.

### ***Participants***

Two groups contributed data for the validation. (1) Operational personnel from seventy-nine (79) OEs volunteered to be interviewed by the JANUS research team. A total of 215 people were interviewed. This group contributed both causal factors data and feedback about the technique. Most were from the radar rather than tower environment. (2) A convenience sample of air traffic personnel in management and staff positions was solicited. The sample averaged 21 years of air traffic experience. This expert forum gave their feedback about the practicality and usability of information derived from the technique.

### ***Method***

Field Interviews. The JANUS taxonomy was scripted into a computer interface and the computer was transported to the facility by the researcher conducting the interviews. A feedback form for participants was also developed.

A team of researchers was trained on the technique for the interview procedure. When an OE occurred and the controller who was working the traffic volunteered to participate a researcher travelled to that facility. All of the OEs analysed by the JANUS team occurred between 12/06/2001-8/07/2002 at 12 air traffic control facilities. Interviews were conducted individually. When a re-creation of the incident was available, each participant was given the opportunity to watch it prior to the interview. Re-creations

were available for 77.7% of the interviews. Feedback forms were left with the participants at the conclusion of their interviews to be filled out and returned. If the participant had been interviewed for another OE, the participant was not given another feedback form.

Forum Feedback. Five incidents from the field test were selected in a quasi-random manner so that no member of the forum would be rating an incident from his or her facility. Data from the field interviews were de-identified and summarized into a format comparable to the current process's tabular format of causal factors classification.

Participants compared information about causal factors developed from the JANUS interviews with outputs from the current process. Each scenario was rated on six dimensions related to the validation criteria: specificity informativeness, comprehensiveness, usefulness, practicality, and time needed to use the information produced from the technique. The comparison was done in a side-by-side manner with pencil and paper using a 10-point scale anchored by Much Less (1) to Much More (10). Multiple OE scenarios were evaluated by each participant. The scale was reversed for the assessment of the time needed to use the information to develop an OE mitigation plan. In this case, a lower scale value (Much Less) indicated greater value for JANUS compared to the current process.

## **Results**

The outputs of the current FAA report (FAA Form 7210-3) was compared to outputs from the JANUS technique. While comparing these, remember that the current FAA technique views the incident overall and analyzes it as a unitary event. Besides several types of descriptive information, the current FAA report identifies causal factors in categories of Data Posting, Radar Display, Aircraft Observation (Towers Only), Communication Error, Coordination, and Position Relief Briefing. The JANUS technique, on the other hand, approaches the incident as potentially having multiple "links

in the chain" and permits analysis of each link separately.

Field Data. Data from 79 OEs were available: 64 from air route traffic control centers (ARTCCs) and 15 from terminals. On the FAA report 7210-3, one hundred thirty three (133) causal factor items were reported for the 79 OEs, an average of 1.7 per OE (range 1-5). Categories of causal factors and the percent represented in this sample of reports were

- Data posting (9.8%)
- Radar display (58.7%)
- Aircraft observation (towers) (1.5)
- Communication error (25.6%)
- Coordination (4.5%)
- Position relief briefing (0%)

JANUS data from 215 interviews with operational personnel was used for the comparison. Interview participants represented several operational roles: the controller working traffic at the time the OE occurred (ATC-1, n=79) and other personnel (non ATC-1, n=136) who could add perspective to the situation. Six types of operational roles were represented in the interviews: 111 operational air traffic control specialists (ATCS) and this group was further broken out as the 79 focal controllers who were working the traffic at the time of the OE (ATC-1) vs. the 32 other ATCS (non ATC-1), such as a handoff controller; 7 controllers-in-charge (CICs) who are those controllers who are qualified to act as supervisors when needed; 3 instructors providing on the job training for the ATC-1 when the OE occurred (OJTI); 61 operational supervisors; 20 operations managers; 12 facility managers; and 1 role identified as "other."

To compare with the 7210-3, the data were first examined for quantity of factors identified and then redundancies were eliminated to identify the unique items within each group. The data were categorized several ways for different purposes based on 79 OEs, 215 interviews, 79 ATC-1 (117 critical points analyzed), and 136 NonATC-1 (198 critical points analyzed).

In interviews with ATC-1 participants the following categories of factors were identified. To illustrate, the category of Perception and Vigilance was reported to be influential in 41% of the critical points.

- Perception & Vigilance 41%
- Memory 15%
- Planning & Decision Making 49%
- Response Execution 10%

ATC-1 interviews identified a total of 281 cognitive factors from the Error Detail categories, an average of 3.6 psychological factors per OE. Of these, 52 were unique concepts, such as “visual search failure.”

Interviews with all participants (315 critical points analyzed) produced a total of 762 contextual factors, an average of 9.6 per OE. The following categories were reported:

- Traffic & Airspace 49%
- Weather 28%
- Teamwork 26%
- Pilot Actions 21%
- Personal Factors 21%
- Pilot-Controller Communications 20%
- Ambient Environment 18%
- Workplace & HMI 13%
- Procedures & Orders 11%
- Training & Experience 10%
- Supervision & Mgmt 10%
- Organizational Factors 10%
- Interpersonal & Social 5%
- Documents & Materials 0.3%

Inter-rater Agreement. Several constraints placed on the field interview process to minimize impact of the research on operations during this beta test made a strict assessment of inter-rater or intra-rater agreement impossible. Absent the rigorous methodology required to assess these, an alternate analysis was conducted to determine whether the technique would be used similarly by different people for the same event.

Data at the Error Detail level from all roles were used to examine agreement between the ATC-1 who was working the

traffic at the time of the OE and the responses by the other participants interviewed for that OE (e.g., the hand-off controller, the supervisor, the operations manager). Although they were more distant from the actual event, all were air traffic control specialists. Those in management roles other than supervisors and controllers-in-charge (CICs) were not required to maintain currency, however. Percent agreement and Cohen’s kappa between the controller working the traffic and other participants are shown in Table 2. The table reflects the unbalanced number of participants by role across the incidents.

**Table 2. Agreement between ATC-1 vs. Other Participants.**

Operational Role	N	% Agree	K	Sig.
OJTI	12	75	--	--
Non ATC-1	164	73.2	.48	.0000
CIC	32	71.9	.46	.0000
Supervisor	328	71.3	.44	.0000
Operation Manager	104	66.3	.33	.0000
Facility Manager	52	65.4	.47	.0000

Sensitivity Comparison. A sensitivity matrix resembling a signal detection matrix [11] was used to compare the causal factors identified by the two techniques. This compared the “hits” and “misses” between the causal factor data reported on the FAA form and in the JANUS categories to determine similarities and differences between them. This analysis approach provided evidence to determine whether the JANUS technique added any value beyond the current process.

An ATC subject matter expert who was familiar with both techniques examined the Causal Factors block on the FAA report and judged 32 items to be causal factors. The remaining 32 items were either descriptive or elaborative elements. The causal factors were then coded according to the JANUS category with which it would be most closely associated. For example, *Failure to Detect Displayed Data* on the FAA report was coded



in the Perception & Vigilance category of JANUS.

- 52% of the items were “hits,” that is, covered by both the 7210-3 and JANUS.
- 0% of the 7210-3 items were “misses,” that is, not covered by JANUS.
- 48% of the JANUS categories were available but went unused.
- The fourth cell of the sensitivity matrix (absent in JANUS and absent in the 7210-3) was empty because existing processes did not provide the missing information.

Participant Feedback. Thirty-three percent of the feedback questionnaires handed out to participants were returned. In general, participants were comfortable about participating in the project (60%), found incident replay to be useful (61%), and at the conclusion of the interview had overall positive opinions about the technique (61%). A majority of participants (56%) thought that the questions asked were relevant to causal factors.

Forum Feedback. Participants compared information derived from the JANUS technique to that of the current process using 10-point scales anchored by Much Less-Much More. Higher scores indicated greater value for JANUS for all but *Time to Use*. In this case, a lower scale value (Much Less) indicated greater value. Results showed a higher rating for JANUS on:

- Comprehensiveness (mean = 7.18)
- Informativeness (mean = 7.18)
- Practicality (mean = 7)
- Specificity (mean = 7.38)
- Usefulness (mean = 7.17)

The forum was ambivalent about how JANUS would compare to the current process on *Time to Use* (mean = 5), probably because they had no experience with the technique to make the comparison. Participants were asked their opinion about the strengths and weaknesses in comparing JANUS vs. the current FAA process. A sample of the written comments is shown in Table 3.

**Table 3. Sample Comments From FAA Forum.**

Strengths of current process
<ul style="list-style-type: none"> <li>• Ease of completion</li> <li>• Less conjecture</li> <li>• Easier to comprehend.</li> <li>• Familiarity</li> </ul>
Strengths of JANUS
<ul style="list-style-type: none"> <li>• Much more info available to identify and fix problems</li> <li>• More detail</li> <li>• Pinpoints the causal factors</li> <li>• More informative</li> </ul>
Weaknesses of current process
<ul style="list-style-type: none"> <li>• Just a check box form</li> <li>• Not any room for leeway</li> <li>• Doesn't get to the problem</li> <li>• Over simplification</li> </ul>
Weaknesses of JANUS
<ul style="list-style-type: none"> <li>• Sometimes too informative and subjective</li> <li>• May not be practical, due to equipment limitations, and staffing.</li> <li>• Identifies too many factors, need to prioritize</li> <li>• Maybe too precise</li> </ul>

## Evaluation of Results

These validation activities were designed to answer five questions:

1. Does JANUS work?
2. How well does JANUS work?
3. Is JANUS better than the current method?

Comparison of the results of these studies against the validation criteria showed that initial analysis of both objective data from interviews and subjective data from the feedback and the forums support the approach.

Taken together, the Eurocontrol and FAA results yield converging evidence that the JANUS technique appears to be more sensitive, useful, comprehensive, and practical than the current processes to identify causal factors.

#### 4. Is JANUS ready for implementation?

These data suggest that the technique has great potential for application, although validity cannot be fully claimed without comparable levels of reliability. While these results support the validity of the technique, some scientific issues remain to be more fully answered through further research before operational implementation. These include (a) identifying improvements to increase agreement and reliability between users, (b) using this information to develop appropriate training for users, (c) refining the taxonomy, (d) further standardization of the methodology, (e) making design changes to the computer-based interface, (f) relating causal factors to objective temporal markers in incidents, and (g) linking JANUS outputs with ATC error mitigation strategies.

#### 5. Will the results from JANUS help to improve safety management?

The results from this project to date appear to affirm this, but a definitive answer will be found after additional data are accumulated and from which information can be drawn to make recommendations for strategies to mitigate the potential for future operational errors. Thus, a more robust determination will be made once we can look back from a longitudinal view.

## References

- [1] EATMP, 2003, *Short report on human performance models and taxonomies of human error in ATM (HERA) – HRS/HSP-002-REP-02*. Brussels: EUROCONTROL.
- [2] Shappell, S.A., & Wiegmann, D. A. (2000). *The Human Factors Analysis and Classification System--HFACS*. (DOT/FAA/AAM-00/07). Washington, DC: Federal Aviation Administration Office of Aviation Medicine.
- [3] Isaac, A., & J. Pounds, 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.
- [4] Pounds, J., & A. Isaac, 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.
- [5] Reber, A. S. 1985, *Dictionary of psychology*. London: Penguin.
- [6] Uebersax, J., 2002, *Statistical methods for rater agreement*. [On-line]. Available: <http://ourworld.compuserve.com/homepages/jsuebersax/agree.htm>.
- [7] EATMP. 2003, *Validation of the human error in ATM (HERA) – HRS/HSP-002-REP-04*. Brussels: EUROCONTROL.
- [8] Fleiss, J. L., 1981, *Statistical methods for rates and proportions*. New York: Wiley.
- [9] Cohen, J., 1960, A coefficient of agreement for nominal scales. *Education and Psychological Measurement*, 20(1). 37-46.
- [10] Kirwan, B., 1992, Human error identification in human reliability assessment. Part 2: Detailed comparison of techniques. *Applied Ergonomics*, 23(6), 371-381.
- [11] Swets, J. A., 1996. *Signal detection theory and ROC analysis in psychology and diagnostics*. Lawrence Erlbaum: Mahwah, NJ.