# Measuring the Effectiveness of Error Investigation and Human Factors Training

***C. G. Drury, Jiao Ma, and K. Woodcock***
*University at Buffalo, State University of New York*
*Department of Industrial Engineering*
*Buffalo, NY 14260*

**May 2002**

**Prepared for:**

**Office of Aviation Medicine (AAM-240) and**
**Flight Standards Service under Contract #DTFA01-94-C-01013**
**Federal Aviation Administration**
**Washington, D.C.**

# Table of Contents

1.0 Executive Summary ..... 3
2.0 Introduction ..... 5
2.1 Significance and Objectives ..... 7
2.2 Research Background ..... 8
2.3 Job Aids in Aviation Maintenance ..... 17
2.3.1 MEDA ..... 18
2.3.2 Five Rules of Causation ..... 20
3.0 Methodology ..... 23
3.1 Incident Scenario Development ..... 24
3.2 Participants ..... 24
3.3 Exerimental Design ..... 24
3.4 Interview Protocol ..... 25
3.5 Analysis Methods ..... 27
4.0 Results ..... 28
4.1 Sample Representativeness ..... 29
4.1.1 Year 1 Versus Year 3 Demographic Differences ..... 30
4.2 Overall Analysis ..... 30
4.2.1 Year 3 Data ..... 31
4.2.2 Year 1 / Year 3 Comparison ..... 35
4.2.3 Analysis of MEDA Facts ..... 35
5.0 Discussion ..... 40
5.1 Do Job Aids Help in Investigation? ..... 41
5.2 How Do Job Aids Help in Investigation? ..... 43
5.3 Availability of Scenario Methodology ..... 46
6.0 Conclusions ..... 46
7.0 References ..... 48
8.0 Acronyms ..... 51

# REPORT

# MEASURING THE EFFECTIVENESS OF ERROR INVESTIGATION AND HUMAN FACTORS TRAINING

## PHASE III

*C. G. Drury, Jiao Ma and K. Woodcock*
*University at Buffalo, State University of New York*
*Department of Industrial Engineering*
*Buffalo, NY 14260*

## 1.0 EXECUTIVE SUMMARY

This report provides the findings from the final year of a three-year study of how effectively aviation maintenance errors and incidents can be investigated. It is important for aviation safety that errors, incidents and accidents be investigated thoroughly to learn the correct lessons to prevent future incidents. While much necessary effort has been focused on analysis of the causes of errors, these analyses ultimately depend for their validity on whether or not the appropriate set of facts was collected by the investigators. During prior years of this project it was established that investigators only collect a fraction of the available facts. The current project was designed to measure the effectiveness of job aids in improving the thoroughness of investigations.

During Year 1 we developed a methodology that involved having participants investigate a known incident scenario by asking the experimenter for facts, as they would in their normal investigation routine. They stopped when they determined that they had sufficient facts for their report to management. The numbers and types of fact requested became measures of how thoroughly each incident scenario was investigated. Scenarios had between 40 and 115 available facts, of which aviation maintenance investigators found 25-35. As job aids for incident investigation and analysis are widely available in the aviation maintenance community, the final year of the project was devoted to measuring whether, and how, such job aids could enhance investigation performance.

The two job aids used were the Maintenance Error Decision Aid (MEDA) developed by Boeing and the Five Principles of Causation developed by D. Marx. Both are used to analyze incidents rather than investigate them specifically, and indeed the Five Principles was not designed as an aid to the investigation process. Because in the prior years no participant had used either aid during our study, we could compare investigators using the job aids in Year 3 with the baseline level of performance established in Year 1. We tested a total of 15 users of the two job aids using the same methodology as previous years, except that the investigators were provided with the job aid they had been trained to use. Eleven of the 15 participants used their job aids during the investigation but four did not.

The results showed a significant improvement in investigation performance when the job aids were used compared to prior years. Three styles of job aid use were found from video-tapes of the participants. Style 1 was the use of the job aid as a checklist, with the investigation broadly following the format and sequence of the job aid. Style 2 was where the job aid was used as a backup tool for an investigation that proceeded along the investigator's chosen line of questioning. Style 3 was where the job aid remained either totally unused or merely filled in as an added task after the investigation. The most important finding was the job aid was only effective when it was used during the investigation. Styles 1 and 2 averaged 56 facts per investigation while Style 3 (including Year 1 Baseline participants) averaged only 33 facts, almost a 70% improvement with

use of the job aids. There were few changes in the types of facts investigated (Task, Operator, Machine, Environment, Social) using the job aids, with generally more of each type investigated when the job aids were used.

It was also possible to perform a different analysis in Year 3 because each of the scenarios was analyzed in detail using the MEDA Results Form. Human Factors practitioners from the project team and Boeing's MEDA group determined which of the facts for each scenario should be included on the MEDA Results Form. These "MEDA Facts" were then analyzed in a manner parallel to the overall facts. The results of this MEDA facts analysis showed that investigators requested a somewhat greater proportion of MEDA facts (41%) than overall facts (33%) across Year 1 and Year 3 data. It was encouraging that investigators were more selective towards the more important facts available to them, even if the effect was not large.

On the basis of this study we conclude that job aids, even if not designed specifically to assist in the investigation stage of incident, are effective if they are actually used. We found almost no specific biases from using the job aids, and would not suggest changes to the current versions of these tools.

The scenario investigation methodology developed in this project is being made available to users either as a training tool or for further experimental evaluations. It provides a unique way to quantify how thoroughly investigators investigate incidents in aviation maintenance.

# 2.0 INTRODUCTION

Aviation Accident Investigation has been recognized by most countries as a necessary component of aviation safety. Many countries and many military services having an equivalent of the National Transportation Safety Board (NTSB) charged with determining the causes of accidents so that preventive measures can be implemented. Researchers and safety practitioners agree that it is also useful to investigate incidents

where the outcomes are less severe than accidents. The usual assumption is that the same causal factors are involved in both accidents and more minor incidents, so that prevention of the more common incidents will help prevent the extremely rare accidents. This assumption has been tested recently in the U.S. Navy (Schmidt, 2000)[1] and found to be realistic. If all agree that incidents should be investigated to reduce accidents, then it behooves the whole aviation community to optimize the process of incident investigation. The current study helps this optimization.

Phases I and II established that the incident investigation methodology we developed was a powerful and sensitive tool for understanding how aviation personnel investigate maintenance incidents. We found that overall only about 25-35 of the available set of facts (out of a maximum of 40-115 facts) were requested by participants. Of this total requested, only about 10 appeared in the participants' synopsis of the incident. There were differences in total facts requested between personnel job types, mainly as a result of including in Phase I a sample on non-airline professional accident investigators, who found about 20% more facts than AMTs, managers or Quality Assurance (QA) investigators.

The main finding of Phase II was that training in Human Factors improved incident investigations. However, it was decided after Phase II that the use of our incident investigation methodology to measure the effectiveness of training programs was perhaps not its most effective use. Hence, in consultation with our FAA project monitor and management team, we changed the direction of the project in this final phase to concentrate on improvement of the incident investigations process. In particular, we focused on the use of job aids, such as incident investigation systems for aviation maintenance, and whether they in fact produced measurable improvements in investigation performance.

## 2.1 SIGNIFICANCE AND OBJECTIVES

Since the focus of the overall project has changed, this section is changed slightly from that given in Phase I (Drury, Richards, Sarac, Shyhalla and Woodcock, 2000)[2] and Phase II (Drury, Richards, Ma and Sarac, 2001).[3]

Continuing error reduction, particularly for human errors, has been a goal of the Gore Commission report (White House Commission on Aviation Safety and Security, 1997)[4] and National Transportation Safety Board (NTSB) directives to the Federal Aviation Administration (FAA). Maintenance errors have been assuming greater prominence over the past several years, and now constitute a major threat to the continuing reduction in accident rates. Within the aircraft maintenance industry, the most common responses to this need have been human factors programs aimed at (a) training Aircraft Maintenance Technicians (AMTs) and others in Maintenance Resource Management (MRM), and (b) implementing human factors-based incident investigation methodologies, such as Boeing's Maintenance Error Decision Aid (MEDA). Both of these programs attempt to change the way AMTs and others approach their jobs by promoting greater understanding of the human factors considerations underlying human work and error causation.

This research program develops and validated in Phase I a common methodology to provide operational measures of relevance to industry for these two, and other interventions. By measuring how well participants investigate incidents, a common measure (number of causal factors identified) is obtained. Also, diagnostic information on what participants still lack in effective use of human factors knowledge can be obtained from the same methodology. These measures were used in Phase II to provide before-and-after evaluations of specific interventions. In Phase III we extend the methodology to evaluate the effectiveness of incident investigation tools and job aids.

Specific objectives for Phase III are:

1. To measure the effectiveness of selected incident investigation methodologies when used in our incident investigation simulation.

2. To suggest improvements to these methodologies to give better performance in incident investigation.

3. To make the methodology available to the industry for future evaluations of incident investigation job aids.

## 2.2 RESEARCH BACKGROUND

The current research program is unique in that it explores the incident investigation process by having participants investigate incident scenarios. Each scenario consists of a relatively exhaustive listing of facts pertaining to the incident. The facts are initially unknown to participants whose task is to elicit facts from the experimenter until they are satisfied that they have satisfactorily investigated the incident. At that point they provide the experimenter with a synopsis of the incident in their own words. Their success is judged primarily by the number and type of facts they elicit and the number and type of facts they choose to include in the synopsis. This methodology was originally developed by Woodcock and Smiley (1999)[5] for analyzing how industrial accident investigators performed their task. In Phase I, we adapted this to the aviation maintenance domain, testing a sample of 37 incident investigation personnel at our airline partners and at a professional investigation service. In Phase II we used this proven methodology to measure the effectiveness of training interventions.

Human error has long been seen as a primary causal factor in accidents, including aviation accidents. Civil aviation has developed an enviable safety record by introducing multiple barriers to the propagation of error through a system (Reason, 1990)[6]. Using

techniques such as redundant inspections, independent inspection of maintenance, automation and a visible paperwork trail, the industry and regulators have helped ensure that a single error (human or other) does not lead to an accident. The accident sequence typically involves multiple events, even though a single event, known as the active failure, is the final unrecoverable point. Before this there have been many conditions, often lying dormant until triggered by an unusual event, which have contributed to the accident. These are known as latent failures or latent pathogens (Reason, 1990)[6]. They can range from adverse environmental conditions, to management and hangar-floor practices. Their common characteristic is that they cause no problem under benign circumstances, but when combined or placed under stress, they emerge as contributing causes of the accident. Human factors engineering begins with the premise that such latent pathogens are inherently predictable from models of human behavior (Hollnagle, 1997)[7], and can thus be designed out of the system.

The genesis of the current project lies in the work of Marx (1998)[8] who studied the causation of accidents using classical attribution theory. He found that people in aviation maintenance have certain consistencies in attribution of incidents, and proposed a set of causation conditions based on these consistencies. However, our point of departure from his work was the assertion that the investigation process itself is an active rather than a passive task, and depends intimately on human cognition. Thus, an investigator must actively choose what lines of investigation to pursue, and when to stop following each causal chain. These decisions are likely to be influenced in a dynamic manner by the number and sequence of facts discovered, as well as by any biases or prejudices of the investigator. Hence, a study of attribution of causes and blame needs to be paralleled by a study of what set of facts an investigator discovers, and what sequence is used to discover them.

In Phase III we are specifically concerned with the use of investigative tools to determine how they affect (hopefully improve) the depth of the investigation. For this we need to understand how people investigate incidents, and how tools designed to improve incident investigation were developed.

The literature on accident, incident and error investigation was originally reviewed in Drury et al (2000).[2] Since then we have performed two phases of the current project in which we have proposed a new model of how people investigate incidents, and have found more literature on the subject. In essence, previous investigators have proposed that an incident investigation has a data collection phase followed by a later data analysis phase. We have found that in fact people investigate incidents by an iterative process in which data analysis leads the investigator to formulate hypotheses about what happened and why, thus leading to further data collection. The process ends when the investigator is satisfied that a satisfactory explanation has been discovered that integrates the known facts.

First, the traditional view of incident investigation is that it consists of four phases:

**Phase 1: Trigger**: An incident will only be investigated if the external results of the incident trigger some action. Given a triggering event, the magnitude of the consequences of the error often determines the way in which the investigation is carried out. Indeed, an influential safety text (Hammer, 1989)[9] states that the exact number and personnel involved in an accident investigation *should* depend on the severity of the injury or damage (page 267), i.e. on the triggering event.

**Phase II: Data Collection**: Typically, investigation procedures are described in legalistic terms rather than human factors terms. Investigators are exhorted to "follow all leads." Most texts (Hammer, 1989,[9] Chapter 15; Ferry, 1981,[10] Chapter 3) concentrate on the physical methods of accident investigation, for example on how to preserve evidence, how to photograph the accident scene, or how to interview accident participants and witnesses.

A major determinant of the quality of the whole investigation process is the depth to which the incident is investigated. Rasmussen (1990)[11] uses the term "stop rule" to describe the decision criterion that the investigator uses to determine when "enough" data

has been collected. At the lower limit, the investigator can stop at the "sharp end" of the incident (Reason, 1990)[6] by determining who was to blame for the event, usually the last people to interact with the system before the incident. This level will usually suffice for legal purposes, for example by finding that a written rule had been violated. But as Maurino, Reason, Johnson and Lee (1995)[12] note, we are increasingly seeking the latent pathogens in our complex and well-defended technological systems. Rasmussen (1990)[11] notes that investigations stop when the analyst finds a sequence of events that matches a familiar prototype. He sees three reasons to stop: where information is missing, where a familiar abnormal event is recognized as a reasonable explanation, or where a cure is available. There is an intimate relationship between the stop rules chosen and the investigator's model of the processes of causation and responsibility for change.

**Phase III: Data Analysis**: Analysis depends upon the investigator's model of the incident generation process. It attempts to find plausible sequences that fit the known facts. Reason's (1990)[6] book uses a classification of errors into slips, lapses, mistakes and violations all based ultimately on an information processing model, such as those of Norman (1980).[13] Maurino et al (1995)[12] proposes a more wide-ranging model covering latent failures as well as local or active failures. Feggetter (1982)[14] proposes a model of the information processing levels with many similarities to that of Rasmussen (1990)[11] and uses it to develop a human factors checklist for aircraft accident investigation. Even in the realm of injury prevention, Engkvist, Hagberg and Wigaeus-Hjelm (1995)[15] use a model based on failure modes to investigate back injury causation with the aid of a checklist. In a post-accident study, Wiegmann and Shappel (1997)[16] used several models of human error to successfully classify about 90% of the error events in a naval aviation accident database.

**Phase IV: Reporting**: Like data collection and analysis, reporting involves a selection of facts as well as the formation of a coherent structure for the data considered. Traditional texts on accident investigation (e.g. Ferry, 1981,[10] Chapter 16) contain direct advice on reporting. The five essential elements of the report are defined by Ferry as facts, analysis, conclusions, recommendations and summary (page 209). *"Facts should*

*be presented in a logical sequence, stressing those which bear on the mishap process and cause of the mishap."* The analysis section is "*a place to order and analyze the facts"* (all from Ferry, 1981,[10] page 209). The primary result of the reporting phase is a reduction in the amount of data made available. Unless the investigation is extensive and highly regulated (e.g. NTSB) then the raw notes and evidence are rarely included in the report. Thus, some active data reduction process is always taking place at the reporting phase.

On the basis of our results from Phases I and II of the current project we have developed a more realistic descriptive model of how people actually investigate incidents. Two of our four Phases (Trigger and Reporting) were well supported by the study. Investigators were constrained to use a Trigger, here the trigger paragraph of the scenario, so that it is hardly likely that our data would not support this phase. The Data Collection and Analysis phases could not be separated in our study, and indeed it is doubtful whether they ever can be in practice. Initial hypotheses are formed, data is collected to test these hypotheses and new analyses performed based on the outcome. This is an iterative process, as indeed it is for all social processes. For example, Fiske and Taylor (1984)[17] discuss information seeking as a social strategy to reduce uncertainty of attribution (pages 33, 107). Thus, if we must describe the process of incident investigation rather than merely prescribe it, we need to remove the temporal distinction between Data Collection and Analysis and treat them as two steps in an iterative loop rather than as distinct phases.

Figure 1 shows our revised model of incident investigation, concentrating on the replacement for the Data Collection and Data Analysis phases originally proposed. After the trigger stage is the exploration of the boundaries of the system under study. This is primarily a temporal exploration, as the spatial boundaries are largely implicit, e.g. the hangar or the departure gate. In this Boundary Stage the investigator extends the information from the Trigger to help structure the rest of the data collection and analysis, so that in one sense this stage provides a logical bridge to the Sequence Stage.

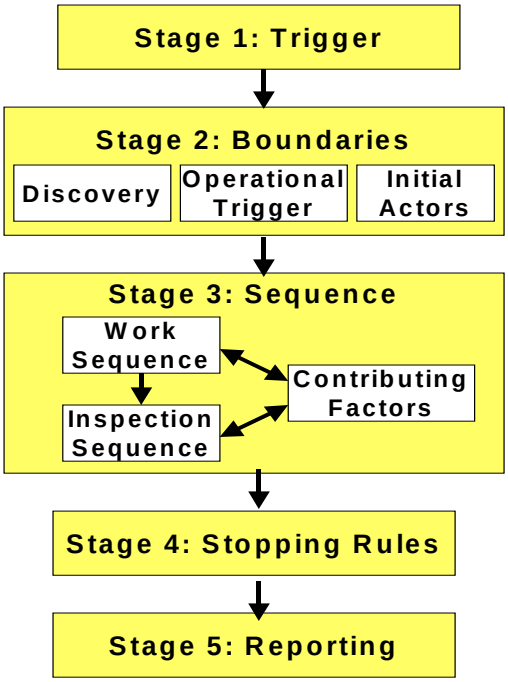


**Figure 1. Model of Aviation Incident Investigation**

The Sequence Stage is where the investigator collects data in a more-or-less chronological order, as we found occurred significantly, starting with Work Sequence and continuing to Inspection Sequence. At times this process must also proceed in reverse as the investigator attempts to reconstruct the incident from both temporal boundaries. In a heavily regulated industry such as aviation maintenance, there are procedural barriers (c.f. Reason, 1997)[18] to error propagation in the form of independent inspection procedures. These are well known to all participants, so that they can deduce that this is a barrier that must logically have failed for the incident to occur. Thus many investigators approached their data collection from both ends, although the temporal ordering predominated.

The Stopping Rule Stage was not requested explicitly. Participants stopped when they had either reached a point that would satisfy their superiors or would be inherently satisfying to themselves. As Rassmussen (1990)[11] put it, they had found a sequence of

events that “matches a familiar prototype”. We found no evidence that would contradict such a statement, but then we did not ask for it.

Our final Stage of Reporting remains from the model originally proposed, perhaps because we asked for a synopsis explicitly. In the synopsis, participants listed the facts and inferences they found most relevant or explanatory. This comprised a much reduced set of facts from that considered in the data collection stages, showing that not all investigation material was considered relevant beyond the investigation. We found no evidence of selection bias, in that some types of fact were suppressed or forgotten. The pattern of fact types retained in the synopsis was almost an exact reflection of that collected in the previous stages.

Overall, the average numbers of facts considered at each stage showed a great expansion from the Trigger stage to the combined Boundary / Sequence / Stopping stages and then a drastic reduction at the Reporting stage (Figure 2).

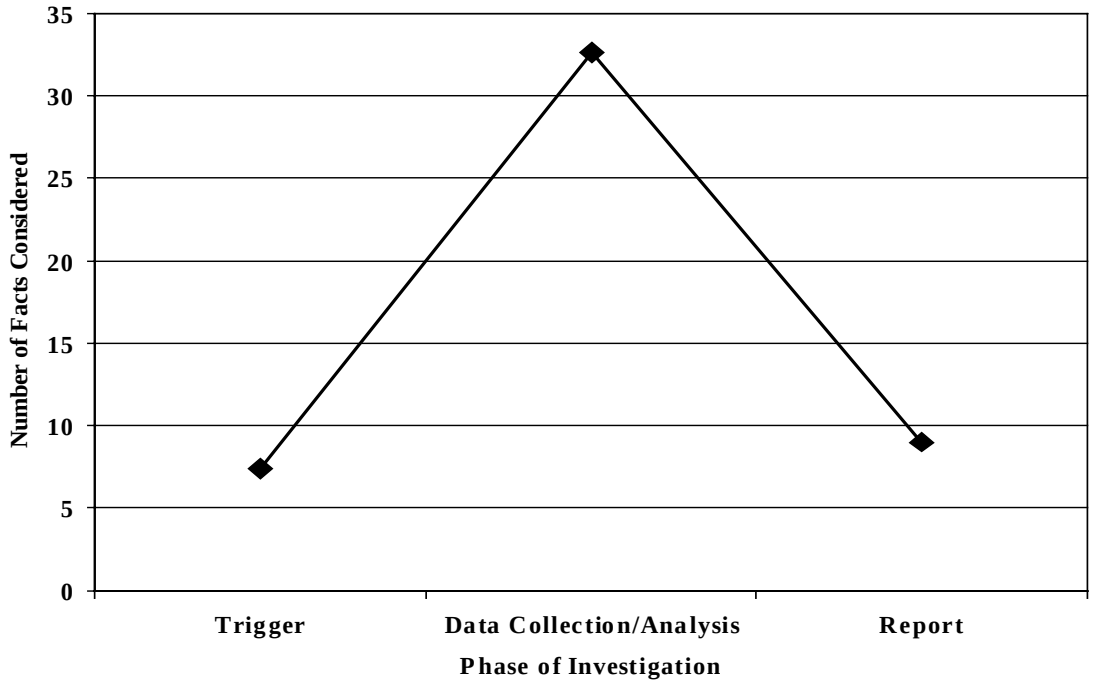


**Figure 2. Data from Year 1 on Facts Considered at Each Stage of Investigation**

Most of the research on incident and accident investigation has concentrated on the analysis of incident data. Codes and taxonomies have been developed and implemented to help understand how and why people have accidents or make errors. Thus, Reason's model sees incidents as instances where behaviors and circumstances have combined to allow an error to propagate through a system, penetrating various safety barriers. This model has been used by a number of organizations as the basis for their investigation procedures, e.g. The Bureau of Air Safety Investigation (BASI) in Australia. O'Hare (2000)[19] extended this framework, using ideas from Helmreich (1990)[20] , Rasmussen (1982),[21] and Wiegman and Shappell (1997)[16] to produce the "Wheel of Misfortune" model. This has local actions (e.g. AMT errors) at its center, surrounded by local precipitating conditions (e.g. resources, task demands) and in turn surrounded by global/ organizational conditions (e.g. culture, philosophy).

There are many other models of the investigation process, but like those noted above, they concentrate on the analysis and reporting stages. For example, Johnson (2000)[22] provides a potentially useful logical symbolism for proving chains of causation using Conclusion, Analysis and Evidence (CAE) diagrams. Interestingly, Lewko (1998)[23] reviews these accident investigation models and the tools derived from them, concluding:

> "One problem with many investigative tools is the tendency to limit the problem solving dimension, thereby restricting the investigator/analyst in developing insights and learning/creating new knowledge. The emergence of these models over time reveal various trends, among which are the gradual shift from searching for a primary cause to multiple causes and the recognition of latent failures as well as active failures."

He notes that "… field investigators have been left to their own devices, or that of external consultants…" leading to prescriptive "menu-driven" methods. Others have considered the differing attributes of blame by victims and their managers (Lehane and Stubbs, 2001)[24] or victims' fallible memory for incidents (Chapman and Underwood, 2000),[25] but again, there is little to guide field investigators. Even reknowned NTSB

managers (eg. Strauch, 1999)[26] concentrate on managerial aspects of scheduling interviews with often distraught participants, or the importance of written reports and public defense of positions, leaving the details of what data to collect to the well-prepared investigation team.

Actual tools for investigation are typically implied by logical analysis of the facts. These can be quite specific to each accident type or domain. For example, Haslam and Bentley (1999)[27] provide a set of interview questions used to investigate slip/trip/fall accidents by postal delivery workers. This uses an accident "sequence of events chart" starting with the immediately preceding activity and proceeding through the fall initiating event to the contact event to the injury event. Each event in this sequence is used to elicit information on environmental conditions, behavior, footware, use of equipment, work methods training and safety practices. A similar, but much expanded, event sequence forms the basis for classifying causal and contributory factors in maritime accidents (Soares, Teixeira and Antao, 2000).[28] While it concentrates on classifying data into a data base, the framework (from the International Marine Organization, IMO) also has direct implications as a job aid for field investigation. A similar investigation interview tool (Sequential Timed Events Plotting: STEP) was used by Green, Morisseau, Seim and Skriver (2000)[29] for investigating incidents on offshore oil drilling rigs. They use a hierarchical questionnaire structure, where a series of initial questions guide the investigator to move detailed questions if required. Rather, in the manner of the Proactive Error Reporting System (PERS) (Drury, Wenner and Murthy, 1997),[30] or the Maintenance Error Decision Aid (MEDA), answers to the initial questions act as filters to guide the investigation down an appropriate path. Hence, not all questions are intended to be asked in every investigation. Finally, the reference to the "Wheel of Misfortune" (O'Hare, 2000)[19] given earlier provides an overall schema of factors to be investigated for their contribution to the incident.

## 2.3 JOB AIDS IN AVIATION MAINTENANCE

Within the aviation maintenance domain, a number of incident investigation methodologies are currently in use. Perhaps the earliest was Boeing's Maintenance Error Decision Aid (MEDA) described more fully below. One of MEDA's developers (D. Marx) went on to produce the Aurora Mishap Management System (Marx, 1998)[31] that expands the concepts introduced in MEDA. Marx then produced a tool that is more an aid to logical reasoning and analysis than a methodology for investigation, the Five Rules of Causation, again described below. In a parallel research program Schmidt, Schmorrow and Hardee (1998)[32] modified an existing human factors investigation methodology, HFACS, for aviation maintenance use in the U.S. Navy. HFACS (Human Factors Analysis and Classification System was originally developed by Shapell and Wiegman (2000)[33] from a taxonomy of possible errors and the causal and contributing factors underlying those errors. As its name implies, HFACS is a system for analysis and classification rather than field investigation, although any well-developed framework can potentially help guide a field investigator towards new and important facts.

Woodcock (1999)[34], who developed and tested the methodology used in the current study, made a telling observation:

> "While there are a number of formalized systems for investigation, safety specialists in the earlier study neither reported using them nor appeared to use them in simulations." From Woodcock (1999),[34] p. 1.

Our own data from Phases I and II of the current study echo this finding. Although many investigators reported training in the use of job aids (such as MEDA) not a single participant in all 71 tests used any form of job aid. A few brought company documentation, such as the MEDA Results Form, to show the experimenters their current tools, but even those did not refer to the tools during the task. Phase III examined specifically the effects of actually using typical tools or job aids during the simulated investigation task.

### *2.3.1 MEDA*

The MEDA process is used by aircraft maintenance organizations to investigate the causes of maintenance errors that lead to safety-related or costly maintenance events, such as flight cancellations, in-flight engine shutdowns, and equipment damage. The MEDA philosophy is that:

- Mechanics do not make errors on purpose.
- Errors are due to contributing factors in the workplace (like poorly written manuals, poor lighting for visual inspection, and not having the correct tool for the job).
- Most of these contributing factors are under management control and can therefore be improved so that future errors are less likely (e.g., rewrite the manual, fix the lighting, and provide the correct tool).

The MEDA investigation consists of an interview with the mechanic(s), who made the error, to understand the contributing factors. A decision is then made by management as to which contributing factors will be improved to reduce future errors. Central to the MEDA process are the MEDA Results Form, currently (2002) in Revision G, and the MEDA Users' Guide (ref).[35] The MEDA Results Form has six sections, moving the investigator from the background information on the incident in a logical manner towards error prevention strategies. Note that a single-incident may trigger more than one MEDA Results Form if more than one error contributed to the incident. MEDA's sections are:

<u>Section I. General Information</u>: background data such as date, time and aircraft details.

<u>Section II. Event</u>: A classification of the event outcome (e.g. operations process event, aircraft damage event, personal injury) plus a short narrative event description.

<u>Section III. Maintenance Error</u>: A classification of the error (e.g. Installation error, servicing error) plus a short narrative description of the error.

Section IV. Contributing Factors Checklist: Here, a large number of contributing factors under 11 categories (e.g. Information, Job/Task, Individual Factors, Environment) are listed exhaustingly. The investigator checks each factor and provides a short narrative description pertinent to the factor.

Section V. Error Prevention Strategies. This section examines the barriers that were breached for the error to have propagated (e.g. Maintenance Policies, Inspection). From these, a list of recommended error prevention strategies is generated, with each keyed to specific contributing factors from Section IV.

Section VI. Summary: A narrative summary of the event, error and contributing factors is required.

In the MEDA Users' Guide, each of these sections is seen in a linear sequence. Sections I through III are characterized as establishing what happened, Section IV establishes why and how, Section V pinpoints failed system barriers and outlines potential solutions. The Users' Guide then goes on to expand each question in detail with examples for each of the Contributing Factors.

Following completion of the MEDA Results form, the data is expected to be entered into a database and feedback then given to all employees affected (e.g. Rankin, 2000).[36] A database is potentially useful to find common causes beyond the resolution of individual incidents (c.f. Drury, Wenner and Kritkausky, 1999),[37] although in practice the analyses of such databases tend to be very one dimensional. Typically the analyses list the top causes, the top aircraft types, the top hangars, etc. without using cross-tabulation to guide more focused interventions.

MEDA was developed by Boeing in conjunction with several airlines, labor unions and the FAA. It is the most widely used aviation maintenance incident investigation tool, with Rankin (2000)[36] reporting implementation in over 120 organizations, and active use

by two-thirds of these. One airline reported decreasing flight departure delays due to mechanical problems by 16%. Another reduced operationally significant events by 48% over two years. Such results can only be achieved if the recommendations made by MEDA are acted upon. More subtly, results can only come from an investigation process that accumulates the appropriate facts. Our current study was a direct test of whether this assertion was true.

### *2.3.2 Five Rules of Causation*

The causation system pioneered by Marx (e.g. Marx and Watson, 1999)[38] was developed to fill a need in incident reporting systems, and particularly the recommendations coming from existing systems. This system is intended to increase the rigor of deriving recommendations from incident data, in a way analogous to Johnson's (2000) CAE diagrams.[22] Note that the Five Rules of Causation were never intended as an investigative job aid, only to assist with making recommendations based on the investigation.

Based on attribution theory (Fiske and Taylor, 1984)[17] and a data collection involving participants who derived attributions from scenario data, Marx originally developed seven causation rules:

1. Causal statements must clearly show the “cause and effect” relationship.
2. Negative descriptors (such as poorly or inadequate) may not be used in causal statements.
3. Each human error must have a preceding cause.
4. Each procedural deviation must have a preceding cause.
5. Failure to act is only causal when there is a pre-existing duty to act.
6. Causal searches must look beyond that which is within the control of the investigator.

7. Statements of culpability must be accompanied by an explanation of the culpable behavior and its link to the undesirable outcome.

These have since been truncated to five rules and taught extensively to airlines, the armed forces and medical practitioners. The five rules are currently stated (from the VA National Center for Patient Safety web site)[39] as:

**Rule 1 - Causal Statements must clearly show the "cause and effect" relationship.**

This is the simplest of the rules. When describing why an event has occurred, you should show the link between your root cause and the bad outcome, and each link should be clear to the RCA Team and others. Focus on showing the link from your root cause to the undesirable patient outcome you are investigating. Even a statement like "resident was fatigued" is deficient without your description of how and why this led to a slip or mistake. The bottom line: the reader needs to understand your logic in linking your causes to the outcome.

**Rule 2 - Negative descriptors (e.g., poorly, inadequate) not used in causal statements.**

As humans, we try to make each job we have as easy as possible. Unfortunately, this human tendency works it way into the documentation process. We may shorten our findings by saying "maintenance manual was poorly written" when we really have a much more detailed explanation in our mind. to force clear cause and effect descriptions (and avoid inflammatory statements), we recommend against the use of any negative descriptor that is merely the placeholder for a more accurate, clear description. Even words like "carelessness" and "complacency" are bad choices much more detailed explanation in our mind. to force clear cause and effect descriptions (and avoid inflammatory statements), we recommend against the use of any negative descriptor that is merely the placeholder for a more accurate, clear description. Even words like "carelessness"

and "complacency" are bad choices because they are broad, negative judgments that do little to describe the actual conditions or behaviors that led to the mishap.

**Rule 3 - Each human error must have a preceding cause.**

Most of our mishaps involve at least one human error. Unfortunately, the discovery that a human has erred does little to aid the prevention process. You must investigate to determine WHY the human error occurred. It can be a system-induced error (e.g., step not included in medical procedure) or an at-risk behavior (doing task by memory, instead of a checklist). For every human error in your causal chain, you must have a corresponding cause. It is the cause of the error, no the error itself, which leads us to productive prevention strategies.

**Rule 4 - Each procedural deviation must have a preceding cause.**

Procedural violations are like errors in that they are not directly manageable. Instead, it is the cause of the procedural violation that we can manage. If a clinician is violating a procedure because it is the local norm, we will have to address the incentives that created the norm. If a technician is missing steps in a procedure because he is not aware of the formal checklist, work on education.

**Rule 5 - Failure to act is only causal when there was a pre-existing duty to act.**

We can all find ways in which investigated mishaps would not have occurred--but this is not the purpose of causal investigation. Instead, we need to find out why this mishap occurred in our system as it is designed today. A doctor's failure to prescribe a medication can only be causal if he was required to prescribe the medication in the first place. The duty to perform may arise from standards and guidelines for practice; or other duties to provide patient care.

These rules are often taught as part of an overall Root Cause Analysis (RCA) program, where the real investigative work is performed prior to the use of the Five Rules in deriving logical statements about causation. In line with our previous research findings,

RCA hints at investigation combining data collection and data analysis. For example, the National Center for Patient Safety uses Root Cause Analysis (RCA) with the Five Principles, and emphasizes the iterative nature of incident investigation with:

> "Remember, doing an RCA is an iterative process. As you learn more about the big picture for an event, use the triage questions and the Five Rules of Causation again to get a clearer idea of what happened in order to prevent a recurrence."

As the data is analyzed, investigator biases can cloud the deductive reasoning process. Marx and Watson (1999)[38] found evidence that positive and negative descriptors in the incident narrative influenced attribution, as did the existence of a rule/procedure violation and the presence of a possible prevention strategy. While these were found to be significant biases in the attribution process, they are also likely to influence the investigation process. For example, evidence of a rule violation may well cause the investigator to terminate the data collection /data analysis stage, believing that the "cause" has been found. This appeared to be true for a number of our participants in previous phases (Drury et al, 2000;[2] Drury et al, 2001[3]). Thus, while the Five Rules of Causation were not intended as an investigative tool, the discipline this system brings to the logical analysis may well be beneficial in incident investigation. This reasoning, plus the widespread use of the Five Rules of Causation in aviation maintenance, argued strongly for its inclusion in the set of tools evaluated in the current study.

# 3.0 METHODOLOGY

The objective of Year 3 was to measure the effects of job aid intervention (e.g., accident investigation tools such as MEDA, Five Rules of Causation) on participants' ability to discover facts in an accident investigation. Based on the results of this demonstration and direct observations in previous years, we will make suggestions on how to improve these interventions, and make our methodology available to the industry in the form of a self-evaluation package. As in Years 1 and 2, we had to develop scenarios, choose subjects, and follow a designed experimental protocol.

## 3.1 INCIDENT SCENARIO DEVELOPMENT

At the end of Year 2, the three scenarios (2,3,4) given in Drury et al. (2001)[3] were reviewed again in light of the analysis and mapping with MEDA categorizations. A small number of changes in TOMES categorization were discussed between the analysts. All of these changes were made before the re-analysis of all Year 2 data reported in the Year 2 report. Therefore, the data produced in Year 3 are directly comparable to those reported in Drury et al (2001).[3]

## 3.2 PARTICIPANTS

In Year 3, we recruited from our partner airlines participants who actually conduct maintenance accidents/incidents investigations at work. Their average experience as an investigator was about 4 years, and they had investigated on average about 16 cases in the previous year. Half of the participants were quality assurance (QA) investigators, and the other half was made up of AMTs and managers. Most of the participants had previous training in both human factors and investigation (e.g., MEDA). Hence, the participants in Year 3 were much more experienced investigators that those in Year 2 but were more equivalent to those of the baseline study.

As in the Years 1 and 2, related experience and demographic data was collected on each participant: age, years as AMT (A&P licensed), years in current job, years as investigator, time since last investigation, number of incidents investigated in the previous year, human factors training, and investigation training.

## 3.3 EXPERIMENTAL DESIGN

In the current study, the objective was to test the effect of using investigation job aids such as the MEDA Results Form and the Technical Mishap Investigation Report based

on the Five Rules of Causation. The participants behaved quite differently toward the job aids we provided. At the first two sites, participants used the job aids extensively, while at the third site they did not refer to the job aids at all.

Each participant was randomly assigned to a scenario. As in the Year 2 study and the baseline experiment, there would generally be unequal numbers of participants in each combination of three sites and three scenarios, i.e. the nine cells of the experiment.

The participants in the current study were treated as using investigation job aids, compared with those participants in the Baseline study who were not provided with any investigation job aid.

## 3.4 INTERVIEW PROTOCOL

This was unchanged from previous years, except that a camcorder instead of a tape recorder was used in order to capture the details of how a participant used the job aids. No facial feature was included in the videotapes, which focused only on participants' hands and forearms. Videotaping started upon completion of collecting demographic data. The following is adapted from Drury et al (2001).[3]

The data collection was in interview format, where the participants asked questions which were answered by the experimenter. The interaction with each participant took place in an enclosed and quite room, usually an office or meeting room, at their own worksite. They were provided with the written briefing and consent form and asked for demographic data and permission of being videotaped through the investigation. Then they were once again verbally assured of anonymity. All data was stored by participant number only.

The job aids were laid out in front of the participants. In addition, participants were given a pad and paper to record facts if they desired. The incident trigger paragraph was given to the participant. At this point, the participant was prompted to ask questions of the

experimenter as if they were asking the same questions of personnel in the incident. Questions could cover the range from “What was the temperature in the repair hangar?” to “Did the mechanic have any personal problems that night?” The participants were observed to ask questions either by simply following the structure of the job aids or by structuring their own questions, using the job aids as an additional reference and reminder.

The experimenter answered the participant’s questions from the data sheets developed for each scenario. If the question was not anticipated, the experimenter replied that no data was available on that issue. This was typical of current practice, where nobody had thought to record, for example, the hangar temperature. If participants repeated a question, this was recorded and the same fact given again.

When the participants declared that they would stop the investigation, they were asked to provide a verbal synopsis of the incident, as they would in writing a report. They were asked to list the contributing factors in their synopsis. For those participants who used the job aids, they always had the investigation forms (e.g., MEDA Results Form) filled out completely before they stopped. Usually, they verbally repeated their summary in “Section VI: Contributing factors/error and event” from the MEDA Results Form.

When they completed the assignment, participants were given a de-briefing. This was to remind them of the purpose of the experiment, to reassure them of anonymity, to provide non-specific encouragement about their performance (if asked), and to remind them not to discuss the specific scenario with colleagues who may become future participants. Their notes on the investigation and the written up in the job aids were collected for later analysis.

## 3.5 ANALYSIS METHODS

(This section modified slightly from Drury et al, 2001.)[3] As each participant's interview was completed, the audiotape, experimenter's interview notes and participant's written notes were reviewed. This allowed a separation of the two parts of each interview: the data collection stage and report stage where a synopsis was given. Each stage was then analyzed in a somewhat different manner.

For the data collection stage, the objective was to determine the number and types of facts requested by the participant. Each fact was noted for the order in which it was requested. The first fact was coded "1", the next "2" and so on. If a participant asked for the same information more than once (a quite common occurrence) only the first request was coded. These data were placed into a MINITAB™ database. Later, each order datum was re-coded as either a "0" for a fact not requested or a "1" for a fact requested. This allowed us to count the number and type of facts requested. The number and percentage of facts requested for each scenario were the primary measure of data collected.

For the report stage, a transcript of each report was made. This allowed the analysis team to make an unambiguous list of the facts incorporated into the final synopsis produced by the participant. The order of these facts was not particularly relevant, as most reports gave the facts in time order rather than the most salient facts early in their list. From the list produced, the total number and percentage of facts were both measured.

The primary statistical analysis was of the following dependent variables:

| | |
|---|---|
| Stages 2, 3 & 4: Data Collection | Number and Percent of total facts requested<br>Number and Percent of each classification of fact requested |
| Stage 5: Report | Number and Percent of total facts reported in synopsis<br>Number and Percent of each classification of fact reported in synopsis |

The data were each analyzed by scenario number (2-4) and participant's position in the organization (AMT, manager/supervisor, QA/engineering). Subsidiary variables such as years of experience, organization, human factors training etc. could be treated as covariates. The main analysis tool was a fixed effects analysis of variance using the General Linear Models (GLM) procedure from the MINITAB™ statistical analysis package. The ANOVA model used was a 3 (Sites) X 3 (scenarios) X 5 (Fact Type, TOMES) fixed effects model with participants nested under groups.

Interanalyst reliability was tested in Year 2 by having both analysts independently review and analyze the facts requested on three scenarios. As one of the same analysts was used in Year 3, interanalyst reliability was not remeasured.

Because in Phase III we had a standard against which to compare the participants' responses, additional analyses were performed using the consensus results from the MEDA analysts to determine which facts *should* have been included.

# 4.0 RESULTS

The first analyses presented are those pertaining to the sample of subjects and their demographics to establish whether they were representative. Next, the overall analyses of variance at the fact-type level are given. Finally, the comparison between participants' results and those of the MEDA experts are given. Note that any comparisons with previous years are made with the Year 1 or Baseline group, as in Year 2 the issue was training so that relatively untrained participants were sought. Year 1 data came from participants who were active in human factors, or incident investigation or both, and thus represents the appropriate comparison group for Phase III.

## 4.1 SAMPLE REPRESENTATIVENESS

A total of 15 participants were tested in this study. They came from three sites, the first using MEDA (6 participants), the second using the Five Principles (5 participants) and the third (4 participants) being trained to use one of these job aids, but in fact choosing not to use the aid during their investigation. There were no differences in age or experience as an AMT between the three sites, but there was a difference between sites in years in current job ($F(2, 11) = 7.04$, $p = 0.011$). The MEDA site averaged 1.8 years, the Five Principles site 4.1 years, but the site where neither was used averaged 14.8 years in their current job. This reflected the fact that all participants at the third site were still practicing AMTs, rather than a mix of AMTs, managers and QA personnel at the other sites.

Our sample was tested against this BLS[39] population data, using a non-parametric Wilcoxon test. This showed that our participants were significantly older (median age = 42.5 years, $p < 0.001$) than the BLS median age of 36.2 years. They were also significantly more experienced (median years as AMT = 16.0, $p < 0.001$) than the BLS[39] population's median experience of 9.4 years. Such a finding has been consistent with many AMT samples used in SUNY's aviation maintenance research over the years. Working with major airlines as partners makes our samples biased towards being older and more experienced, as AMTs often move through career paths in general and regional aviation on their way to the major air carriers. There were no significant differences between the three job types of AMT, manager and QA on either age or years as AMT. However, years of tenure in the same job did show a significant difference between job types ($F(1, 11) = 4.83$, $p = 0.050$). AMTs had much higher job tenure (12.8 years) than either managers (1.0 years) or QA personnel (1.7 years).

### *4.1.1 Year 1 Versus Year 3 Demographic Differences*

The baseline sample consisted of 37 participants, mainly quality assurance and other investigators, with some managers and AMTs who had received investigation training. They can be compared directly to our current sample of 15 participants in terms of their demographics. Table 1 shows these comparisons using a t-test. The two groups did not differ significantly in age, experience in aviation maintenance or job tenure.

**Table 1. Age & Experience Differences between Year 1 and Year 2 Samples**

| Measure | Year 1 Mean (SD) | Year 3 Mean (SD) | t-statistic (significance) |
|---|---|---|---|
| Age | 44.1 (9.2) | 42.5 (6.1) | 0.73(not significant) |
| Years as AMT | 13.0 (12.2) | 15.8 (7.0) | 1.04(not significant) |
| Years in current job | 5.11 (6.5) | 6.0 (6.1) | 0.44(not significant) |

## 4.2 OVERALL ANALYSIS

In our interviews with the participants, we provided investigation job aids, e.g., MEDA Results Form and Five Rules of Causation Summary Form. The participants were told to conduct the investigation and use the job aids as they normally would at their work. Three different styles of using the job aids were observed:

**Style 1: Job Aid as Checklist** Two participants in the MEDA group relied on the job aids extensively. For example, they went through each item on the MEDA Results Form in more or less the given order, and asked relevant questions based on those items. They quoted the phrases or read aloud the contents in the form. They did not take many additional notes during the investigation. These participants went back and forth between sections in the job aids to build up a logical connection between them. They usually stopped after they had completed the whole job aid. MEDA participants completed their investigation after filling out “Section V: Error Prevention Strategies, B: List recommendations for error

prevention strategies". They then gave a verbal synopsis based on their written summary of recommendations.

**Style 2. Job Aid as Back Up** Three MEDA investigators and three Five Rules investigators were observed to first conduct their own investigation independently of the job aid, taking extensive notes. At a certain point of the investigation, when they had apparently asked all the questions they could, these investigators started referring to the job aids. They compared the items in the job aids with their previous notes, and started asking additional questions and taking additional notes. At the same time, they filled out the job aids. They usually stopped when they came to their own conclusions. Then they either gave verbal synopsis first then wrote down the summary in the job aids, or vice versa.

**Style 3 Job Aid Rarely Used** One MEDA and three Five Rules investigators conducted the investigations completely independent of using the job aids. They structured their own investigation while taking extensive notes. They usually stopped when they had collected what the felt to be enough information and gave a verbal synopsis. The one MEDA investigator started filling out the job aid but did not ask any further questions, while the others using this style continued ignoring the job aids until they were debriefed and excused from the interviews.

These observations were not analyzed further, but are used later as a partial explanation of some of the results we found.

### *4.2.1 Year 3 Data*

Analyses of Variance (ANOVAs) of Number of Facts Requested, Percentage of Facts Requested, Number of Facts Requested in Synopsis and Percentage of Facts Requested in Synopsis were performed as indicated in Section 3.6. Only the analyses of Numbers of Facts, and not Percentage of Facts will be presented here as the results were quite similar. First, a correlation analysis was performed of all of the demographic and performance variables for each participant. This showed that the only variable significantly associated with any of the performance measures was Number of Incidents Investigated in the previous year. The correlation with Number of Facts Requested was 0.645 ($p = 0.009$)

and a graph of the two variables, with a regression line for illustration, is shown as Figure 3. It is clear from Figure 3 that participants with greater familiarity with incident investigation request more facts in our study. Number of Incidents Investigated was thus used as a covariate in our analyses, making these Analyses of Covariance (ANCOVAs).

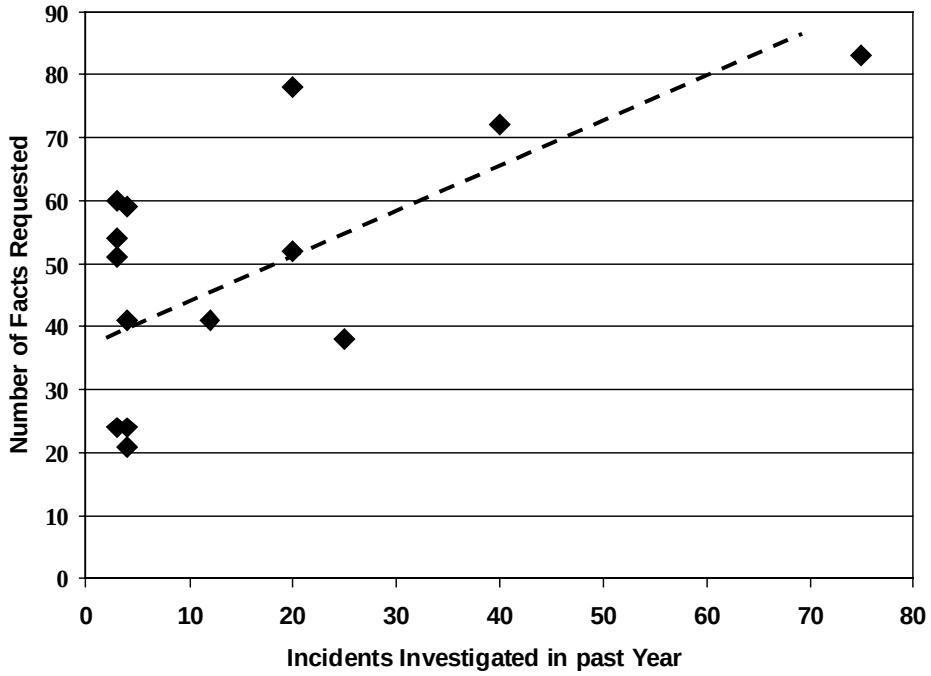


**Figure 3. Correlation between Number of Facts Requested and Number of Incidents Each Participant Had Investigated in the Previous Year**

First, it should be noted that Site did not turn out to be significant for any of the analyses, despite the different job aids used and the lack of real time use by participants at site 3. This was primarily because the variability between the participants was so high, so that even quite large differences in mean facts requested (ranging from 36 to 60 facts) did not reach significance.

For Number of Facts Requested, the covariate was marginally insignificant ($F(1, 27) = 4.10$, $p = 0.052$) but was marginally significant for Percentage of Facts Requested ($F(1, 29) = 4.27$, $p = 0.047$) and so is included in these results. There was a significant effect of Fact Type (TOMES) ($F(4, 29) = 15.91$, $p < 0.001$) and a significant interaction

between TOMES and Scenario ($F(8, 29) = 3.40$, $p = 0.007$). Such effects and interactions have been found in previous years, and indicate that not all fact types were investigated equally. The pattern of TOMES facts requested was again different for the three scenarios. For the Synopsis, the number of facts was significantly different by Fact Type (TOMES) ($F(4, 29) = 5.10$, $p = 0.003$), but not really significant in its interaction with Scenario ($F(8, 29) = 1.97$, $p = 0.087$). The graphs of these two variables (Facts and Synopsis Facts) are shown as Figures 4 and 5 using stacked bar graphs as in previous reports (e.g. Drury et al, 2001).[3]

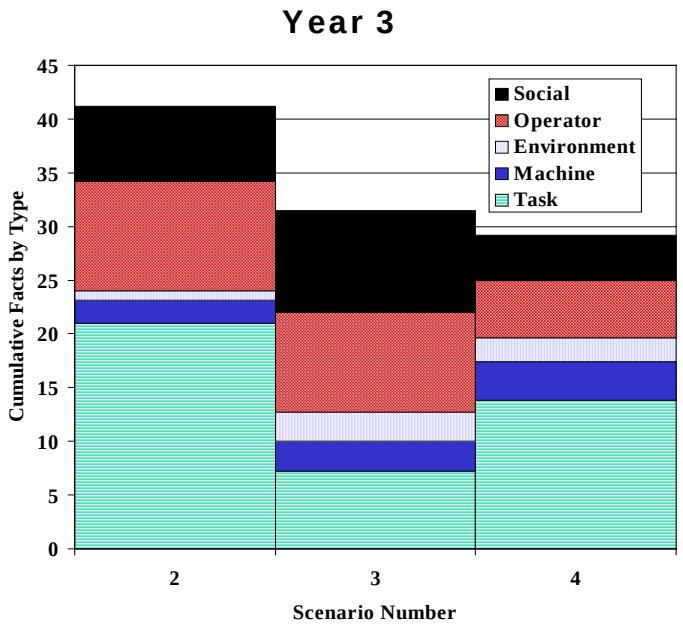


**Figure 4. Year 3 Facts Requested by Fact type and Scenario**

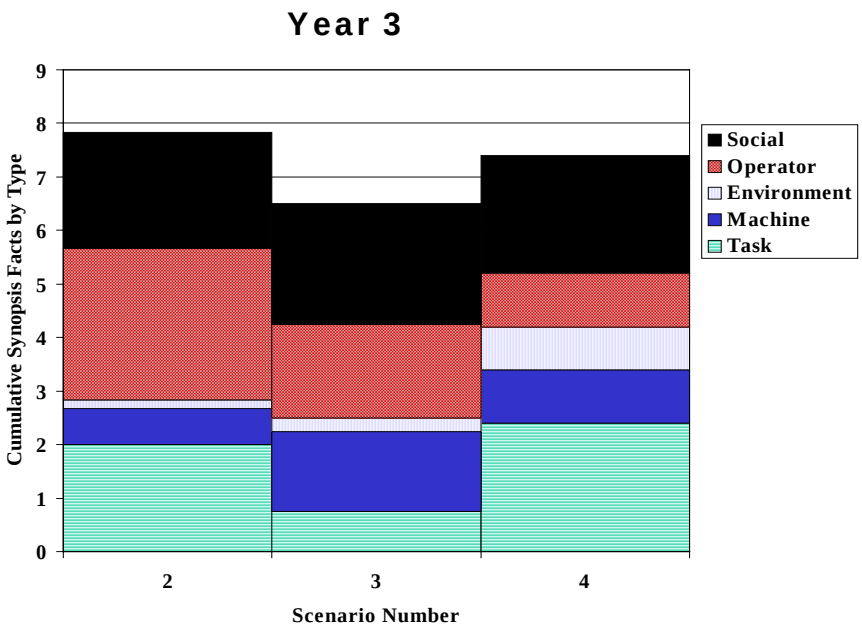


**Figure 5. Year 3 Facts in Synopsis by Fact Type and Scenario**

We will later perform a statistical comparison of this data with the equivalent from Year 1 (Section 4.2.2), but can note two results here. First, the overall numbers of facts in the investigation and synopsis were generally much higher than in Year 1. Second, the pattern of fact types requested and reported across the three scenarios was quite similar to those found earlier. Task facts were still the major contributor, with Operator and Social facts close behind. Fewer machine facts were requested, although a greater fraction of these appeared in the synopsis. Finally, Environment facts were requested and reported rarely, especially for Scenario 2.

It is also of interest to check the times taken to complete the investigation, although accuracy rather than speed is our primary concern. ANOVAs were run of the Year 3 data using Stop Time and Synopsis Time as dependent variables and Site and Scenario as crossed factors. The only significant result was that for Stop Time, Site had a significant effect ($F(2, 4) = 9.31$, $p = 0.031$). The MEDA site averaged 51 minutes for the investigation task, the Five Principles site averaged 35 min, while the third site averaged

only 19 min. Thus, use of the job aids led to different times, with MEDA taking longest and the site not using either job aid taking the least time. In fact, there was a direct, if not quite significant, correlation ($r = 0.505$, $p = 0.078$) between Stop Time and Number of Facts Requested, showing that one interpretation of the job aids is that they force the investigators to consider more facts, thus consuming more time. Across the three Sites, the regression of Facts and Times ($r^2 = 0.943$, $p = 0.154$), was not significant with only three points, but still interesting.

### *4.2.2 Year 1 / Year 3 Comparison*

Because the same type of participants were used in this Year 3 study as had been used in the Year 1 baseline study, a direct comparison of the results between these two years can provide direct evidence of the efficacy of job aids for incident investigation.

Year 1 data were selected for only those participants who were tested with Scenarios 2, 3 and 4, the ones used in the current study. This gave 20 participants from Year 1 out of the total of 37 who received all scenarios. Analyses of Variance were performed for the combined data set with the factors of Job Aid used, Scenario and Job Type. The Job Aid used in Year 1 was in fact no job aid. In Year 2 it could either be MEDA (Site 1), Five Principles (Site 2) or no job aid for Site 3. Again, numbers of facts were analyzed rather than percentage as the analyses gave closely comparable results.

The only significant results were for Number of Facts Requested and Stop Time, where Job Aid was a significant factor. The significance was $F(3, 19) = 5.28$, $p = 0.008$, for Number of Facts Requested and $F(3, 16) = 3.32$, $p = 0.047$ for Stop Time. Post hoc analyses using Tukey Simultaneous Tests at alpha = 0.05 overall showed that for each measure the significant difference was between the MEDA group and the Year 1 group. Both measures are shown plotted against each other in Figure 6, with a regression line added which again failed to reach significance ($r^2 = 0.833$, $p = 0.088$).

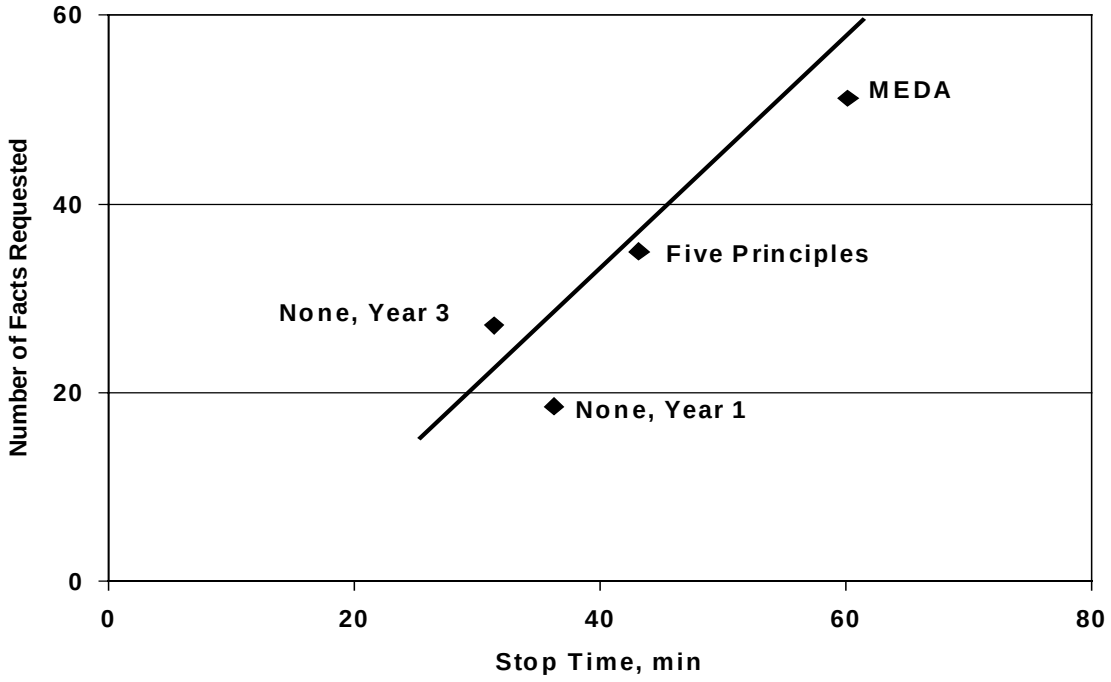


**Figure 6. Relationship between Investigation Stopping Time and Number of Facts Requested for the Different Job Aids**

The overall picture is that the depth of investigation increased with the use of job aids compared with Year 1, and with Site 3 in the current study, although only the MEDA job aid gave a statistically significant improvement.

It is possible to test for Job Aid differences at a deeper level, i.e. whether the job aids had a differential effect on the different fact types requested. The Number of Facts Requested were split by fact type (TOMES) and a second set of analyses of variance were performed, using Job Aids, Scenarios, TOMES and Job Type as the factors. As in previous analyses, Number of Facts Requested was significantly different for Scenario ($F(2, 107) = 2.13$, $p = 0.045$), TOMES ($F(4, 107) = 24.12$, $p < 0.001$) and their interaction ($F(8, 107) = 5.65$, p . 0.001). There was no main effect of Job Aid in this analysis, nor any interactions of Job Aid with TOMES. Thus, the pattern of fact types requested was

not changed by the use of Job Aids, as suggested earlier by the similarity of Year 1 and Year 3 results in Figures 4 and 5.

### *4.2.3 Analysis of MEDA Facts*

So far, as in all three years of this project, we have equated depth of investigation with the number of facts requested, so that more facts requested was taken as evidence of better performance. But not all facts are equally valuable in an incident investigation. The temperature in the hangar was largely irrelevant in the scenario where the cockpit door was not reinstalled, whereas the fact that there was no written instruction to remove the cockpit door was one of the key facts in understanding that incident. Fortunately, the MEDA process can be used as one way to determine which facts in each scenario are the most relevant ones. One definition of a "relevant" fact is when one expert MEDA user fills in the MEDA Results Form and includes a particular fact. It is not of course the only definition: MEDA may give misleading or incomplete analysis, or may lead investigators to less important facts. However, the appearance of a fact in the MEDA Results Form does give a consistent definition, and because MEDA has been used successfully in many airline maintenance organizations, it does have some degree of validity. MEDA was developed by human factors practitioners in aviation maintenance and so has content validity. In this section we show how the MEDA Results Form was used as one example of a "gold standard" to go beyond the counting of all facts as equal.

Our process for defining expert MEDA judgment was for the principal investigator and the two current Boeing human factors engineers responsible for MEDA to fill in MEDA Results Forms on each of the six incident scenarios, then to meet and reconcile any differences. We also obtained the independent advice of a non-aviation specialist in incident investigation (also a human factors engineer). The team considered this other advice and used it as appropriate, but where there were differences in interpretation, we gave less weight to the non-aviation expert. Our research assistant acted as record keeper and analyst in the comparisons of the various choices made by the team. Our reason for using Boeing specialists in this process was that they teach MEDA courses to the aviation industry on a regular basis, including running exercises involving trainees using the

MEDA Results Form to analyze written narrative descriptions of incidents. The rationale for using the principal investigator was that he had been part of the selection, development and analysis of all the scenarios over a three-year period and was intimately familiar with their content. He is also a qualified human factors engineer.

With the consensus thus achieved, we made a separate list for each participant of the facts requested and in the synopsis that were also in the consensus on the MEDA Results Form. We shall refer to these for convenience as "MEDA Facts". We then repeated the analysis of Year 1 and Year 3 data using these MEDA facts in place of the earlier Number of Facts Requested etc. This gave a reduced set of facts that our expert process agreed were important and should be captured by the MEDA process.

The main analysis used Site Scenario and Fact Type (TOMES) as factors. For MEDA Facts, there were significant effects of TOMES ($F(4, 125) = 15.96$, $p < 0.001$) and Scenario x TOMES ($F(8, 125) = 3.96$, $p < 0.001$), as had been found for most previous analyses. Not all Fact Types are investigated equally, and not equally across the three scenarios. The effect of Site did not quite reach significance ($F(3, 125) = 2.20$, $p = 0.091$). For MEDA Facts in the Synopsis, the same factors were significant: ($F(4, 125) = 4.33$, $p = 0.003$) for Fact Type (TOMES) and ($F(8, 125) = 6.53$, $p < 0.001$) for Scenario x TOMES. In addition, Scenario was significant ($F(2, 125) = 3.37$, $p = 0.038$), as was Site x TOMES ($F(12, 125) = 2.32$, $p = 0.011$).

The pattern of MEDA Facts across the scenarios is shown in Figures 7 and 8 in the same form as the overall Number of Facts Requested in Figures 4 and 5. The overall pattern looks similar, except that there are an average of 10.8 facts for MEDA against an average of 36.5 for the comparable overall data. In fact, when these numbers of facts are divided by the maximum facts (either MR+EDA facts or overall facts) the MEDA analysis shows that participants found an average of 41% of MEDA facts but only 34% of overall facts. Clearly, our participants were selective in the facts they requested, finding about a fifth more of the facts judged by our experts to be important. Apart from the overall level of facts reported, the pattern across scenarios and TOMES is not too different. Relatively

less Task facts came from the MEDA data, presumably because the initial part of the MEDA Results Form specifies the most important Task facts, such as date and aircraft tail number. MEDA facts were relatively over represented for Operator and Social fact types, and to a lesser extent for Machine and Environment facts. This shows that the job aids made the investigators relatively more aware of potential contributing factors of importance. In the synopsis (Figure 8) Task facts were relatively few compared to the overall data (Figure 5). Scenario 3 had remarkably few facts of any type reported in the synopsis. The selection of MEDA facts, judged to be the most important, did change the findings although the overall pattern was still recognizable.

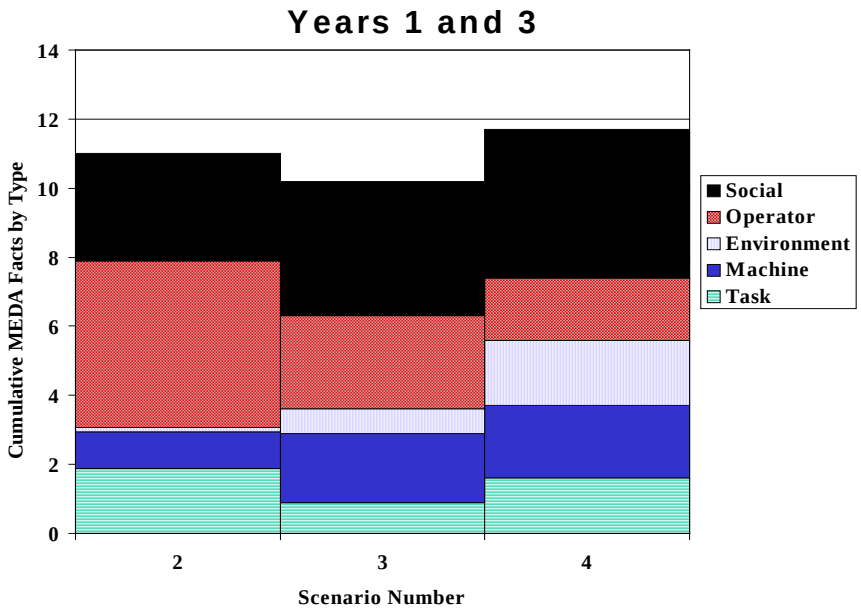


**Figure 7. Number of MEDA Facts Requested by Fact Type and Scenario**

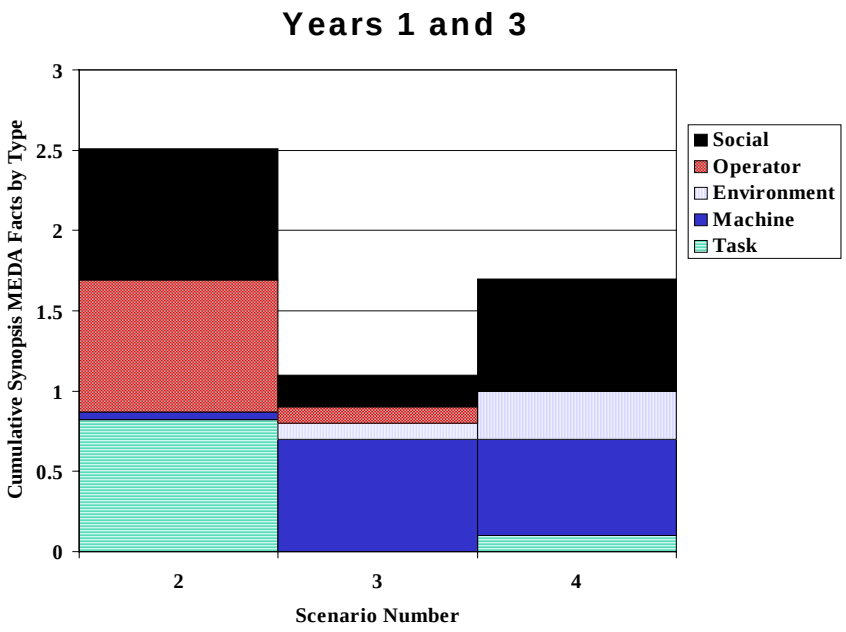


**Figure 8. Number of MEDA Facts in Synopsis by Fact Type and Scenario**

# 5.0 DISCUSSION

We established in prior years that thorough incident investigation was a necessary part of any effort to reduce aviation maintenance errors. We also showed the utility of our methodology of having participants investigate an incident by asking the experimenter questions about a known incident scenario. Using this methodology we established that people only request about a third of the facts available in the scenario, and that the depth depends on the job they are doing. Professional investigators request more facts than AMTs, managers or QA personnel, and training in human factors does help the depth of investigation. Finally, we established a model of the investigation process (Figure 1) emphasizing the close and recursive relationship between discovering new facts and fitting them into a satisfying framework. These phases of data collection and analysis had usually been seen as separate in previous literature.

From this third Phase of the study we have been able to show that having investigators use a job aid enhanced their investigation performance significantly. We now put this finding into the context of both aviation maintenance incident investigation and the literature on incident investigation in general.

To summarize our results:

1. When the either job aid was used, significantly more facts were investigated, although more time was taken in the investigation. This greater depth of investigation applied across our Year 3 participants, but was only significant when we combined the data with Year 1, which used similar investigation personnel.
2. The pattern of Fact Types investigated (TOMES) changed little with the use of job aids, with a similar pattern across scenarios to that found in previous years.
3. Investigators who had investigated more accidents in the past year requested more facts in this study.
4. Facts agreed to be required on the MEDA form were investigated at a somewhat higher percentage than facts overall. The pattern of these "MEDA facts" investigated across fact types and scenarios showed that less Task facts and more of the other types were requested.
5. There were three styles of using the job aids: as a checklist, as a backup or not at all until after the investigation. Those who did not use the job aids provided had about the same performance as our Year 1 participants, who also did not use job aids.

## 5.1 DO JOB AIDS HELP IN INVESTIGATION?

From our model in Figure 1, we can see a rationale for the active use of job aids during investigation. First, such an aid will help remind the investigators to collect the Boundaries data for Stage 2, such as the discovery of the incident, names of those

involved and the event outcome that triggered the investigation. All are needed on most report forms, so that the reminder can be useful, although in practice these facts comprise the initial trigger for many incidents. For example, our trigger for Scenario 2 was:

> **2. Missing Cockpit Door**
>
> During the preflight check on A/C #6833, Flight #1141, the crew found that there was no cockpit door in place. The cockpit door had been removed and not reinstalled during overnight maintenance to locate an under-floor leak.

Following Stage 2, investigators typically went through the event sequence, starting at the work sequence and moving to the inspection sequence, looking for errors in either or usually both. Both job aids encouraged logical thinking, e.g. the classification of the Event and Error in MEDA, and so would support such a process. MEDA listed Contributory Factors in Section IV, giving a hierarchical list rather than a chronological list, which would not support sequence following *per se*. However, such a list would provide a direct *aide memoire* for the Contributing Factors of Stage 3. Stopping Rules (Stage 4) are not given explicitly in either job aid tested, but both require a degree of completeness in turning the contributing factors into a logical argument for the recommendations made. Clearly too, Stage 5 (Reporting) receives much direct assistance from either job aid.

The job aids would be expected to improve performance, even though one (the Five Principles) was never intended as a job aid during the investigation process itself. Indeed, they did make such an improvement overall. We found a significant effect of Site, where different sites used the different job aids, with the whole of Year 1 being classified as a single site. In fact, we can redo the analysis to make this point even more strongly. If we classify Site 3, where the participants did not use the job aid provided, as the same as Year 1, we get a highly significant effect of Site ($F(2,22) = 8.79$, $p = 0.002$). Post-hoc comparisons using Tukey's Simultaneous Tests at $p = 0.05$ showed that the difference between the MEDA site and the others was highly significant. Clearly, job aids are effective, but only if they are actually used during the investigation.

## 5.2 HOW DO JOB AIDS HELP IN INVESTIGATION?

Our videotaping of the use of job aids during the experiment proved most useful in interpreting the way in which the job aids were employed. When we classified the Number Of Facts Requested by Style, we also got highly significant results. We could not perform a multi-factor ANOVA as there were so few of Style 1 and Style 2 participants, so we merely extracted the effects of Style and Scenario. The effect of Style was highly significant ($F(2,30) = 7.68$, $p = 0.002$). Our Style 1 participants, who worked systematically through the job aid, and Style 2 participants who used the job aid as a back up, requested significantly more facts than the rest who were in Style 3. Figure 9 shows this effect clearly.

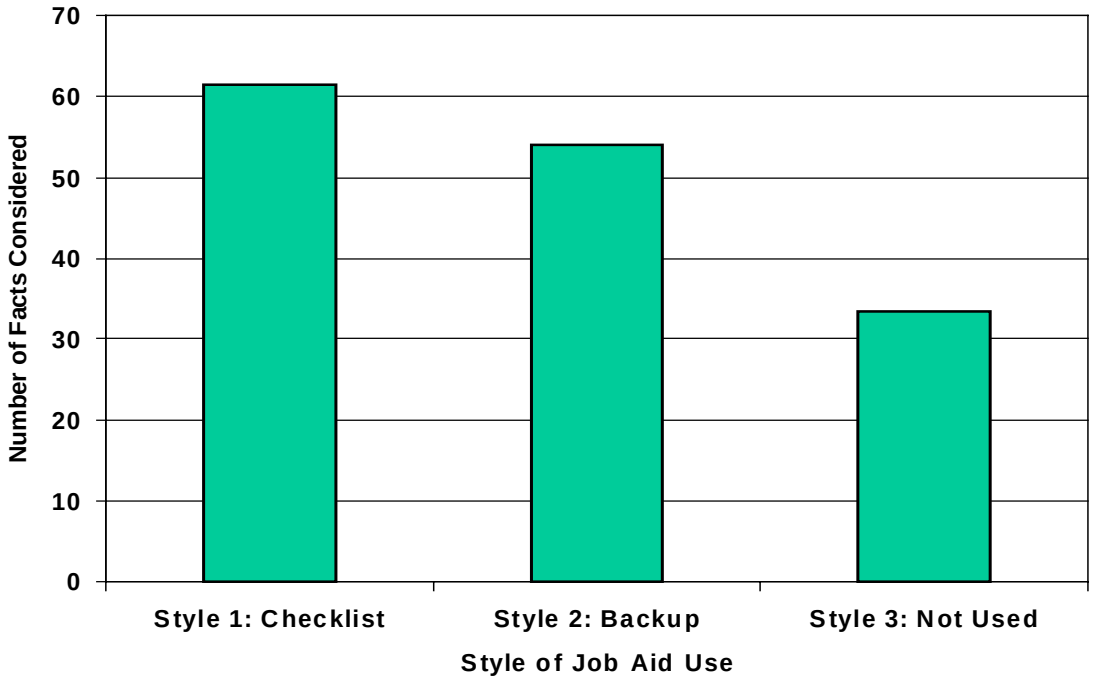


**Figure 9. Effect of Job Aid Use Style on Number of Facts Requested**

We may not have found much difference between the two job aids in terms of performance, but we found highly significant differences between whether they were used or not. As we noted earlier, Woodcock and Smiley (1999)[5] reported that investigation systems, while available, tended not to be used in practice. We too had seen this behavior: in our previous 71 trials across two years nobody used a job aid even when available. Incident reporting systems seem to be regarded as aids to analysis, and particularly reporting of incidents rather than their investigation. In fact, practitioners often use them as ways to populate a data base in a consistent manner, typically with little thought to the analysis of this valuable accumulated resource (Wenner and Drury, 2000).[41] What we have shown in the current study is that the job aids need to be actively used to be truly effective. For example, the training in job aids should include active investigation with the job aid, perhaps as a with-and-without comparison, as well as being used to code a fully written incident narrative. The scenarios developed in the current project could be useful prototypes for such a training program.

In this study we were able to go beyond the previous years and analyze not just the overall facts requested, but one version of the facts that *should* have been requested. Our consensus between the MEDA human factors personnel and project staff on what facts should be included in a MEDA analysis was encouraging, even though our non-aviation human factors practitioner gave a somewhat different interpretation. The analyses of "MEDA Facts" closely paralleled the overall analysis of facts requested, although there were only about one third as many facts fitting this more restricted definition. Investigators found about 41% of MEDA facts as compared to 34% of overall facts which is encouraging. Clearly investigators have some bias towards the more important facts, at least when importance is judged by human factors practitioners.

To a small extent the use of the job aids changes the types of facts requested. We already found that for most of our analyses over the three years, the pattern of fact types investigated was similar, but in one analysis (for MEDA Synopsis Facts) there was a significant Site x Fact Type interaction. Again we can combine Site 3 with Year 1 and get an even more significant result ($F(8,140) = 2.84$, $p = 0.006$). The pattern of MEDA

facts in the synopsis is shown for these three reduced groupings in Figure 10. We have already noted that there are fewer facts in general requested for non-users of job aids, and this is a salient point in Figure 10. Within the three Job Aid categories, MEDA users report more Social facts and relatively less Task facts, while Five Rules users appear to give more weight in their reports to Operator facts. There are three groups of Contributing Factors in MEDA that were classified here as Social: Organizational, Leadership/Supervision and Communication, so the MEDA emphasis on these facts types is to be expected. The Five Rules philosophy does include many examples of individual factors, as shown in our summary in Section 2.3.2, e.g. "resident was fatigued", so again its association with Operator factors is perhaps explainable. Note however, that MEDA synopsis facts was the only analysis to find a Site x Fact Type interaction. The results in Figure 10 are in a minority: for the most part the job aids increased the depth of investigation overall rather than emphasizing any particular fact type. These job aids were relatively unbiased in their application.

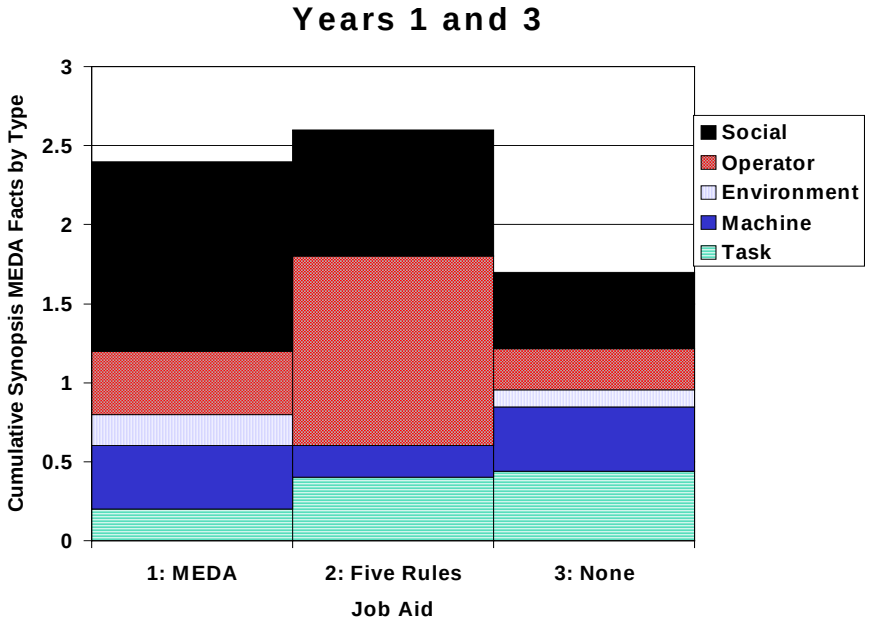


**Figure 10. Effect of Job Aid on Type of Fact Requested**

From our data we did not find a pattern specific facts missed by the investigators, so that there are no specific recommendations for changing the tools. During the MEDA analysis, it was noted that there was really no place on the MEDA form for the human information processing errors traditionally studied by human factors engineers. These would include memory errors, using the wrong control, or biases in decision making. It was concluded that these were not of great value to the typical users of these job aids, who have been trained to interpret human factors in error causation more broadly, e.g. in terms of behavior norms.

## 5.3 AVAILABILITY OF SCENARIO METHODOLOGY

The intention of this project has always been to develop the scenario investigation methodology as a tool for future use by other investigators. We have used it now in three studies and found it to be a valid and sensitive way to measure quantitatively the thoroughness of investigation. We have demonstrated repeatedly that people need help in uncovering facts during an investigation, not just in analyzing the facts into a coherent set of findings. If the facts are not collected in the first place, any analysis must by incomplete at best and biased at worst. We offer the scenarios to any legitimate user, for future studies of investigation, or as a training tool for investigators. We would rather they not become open property, e.g. on a www site, as that could compromise their future use. Please contact the Principal Investigator at drury@buffalo.edu for the scenarios and discussion of how to perform analyses of the results. The basic techniques for their use in research have been covered in previous reports in this series (Drury et al, 2000, 2001).

## 6.0 CONCLUSIONS

1. Job aids such as MEDA or the Five Principles of Causation, can increase the thoroughness of incident investigation.

2. Job aids must be used during the investigation, rather than as post-investigation reporting devices, to be effective.
3. Job aids are equally effective if they are used as a checklist or used as an additional *aide memoire* during the investigation.
4. The two job aids tested did not appear to change the type of facts investigated appreciably, leading to the conclusion that they increase the depth of investigation without biasing the results.
5. The model we have developed in previous reports was upheld in the current study.
6. Our investigation methodology and scenario materials will be made available as research tools or training aids to legitimate users.

# 7.0 REFERENCES


1. Schmidt, J. K., Schmorrow, D. and Figlock, R. (2000), Human factors analysis of Naval Aviation Maintenance Related Mishaps, *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society*, July 20-August 4, 2000, San Diego, CA.
2. Drury, C. G., Richards, I., Sarac, A., Shayhalla, K. and Woodcock, K. (2000). *Measuring the Effectiveness of Error Investigation and Human Factors Training*, Washington, D.C.: Office of Aviation Medicine, FAA.
3. Drury, C. G., Richards, I., Ma, J. and Sarac, A. (2001). *Measuring the Effectiveness of Error Investigation and Human Factors Training*, Washington, D.C.: Office of Aviation Medicine, FAA.
4. White House Commission on Aviation Safety and Security (1997). *Gore Commission Report*.
5. Woodcock, K. and Smiley, A. (1999). Developing simulated investigations for occupational accident investigation studies, prepublication draft.
6. Reason, J. (1990). *Human Error*. Cambridge, U.K.: Cambridge University Press.
7. Hollnagle, E. (1997). *CREAM-Cognitive Reliability and Error Analysis Method*, New York: Elsevier Science.
8. Marx, D. (1998). Discipline and the blame-free culture. *Proceedings of the 12th Symposium on Human Factors in Aviation Maintenance*. London, England: CAA, 31-36.
9. Hammer, W. (1989). *Occupational Safety Management and Engineering*. 4th Edition, NJ: Prentice Hall.
10. Ferry, T. S. (1981). *Modern Accident Investigation and Analysis*. NY: John Wiley and Sons.
11. Rasmussen, J. (1990). The role of error in organizing behaviour. *Ergonomics*. **33.10**, 1185-1199.
12. Maurino, D. E., Reason, J., Johnston, N., and Lee, R. B. (1995*). Beyond Aviation Human Factors.* Aldershot. England: Avebury Aviation.
13. Norman, D. A. (1980). *Errors in Human Performance*. Report #8004, Center for Human Information Processing. San Diego: University of California.
14. Feggetter, A. J. (1982). A method for investigating human factor aspects of aircraft accidents and incidents. *Ergonomics*, **25.11**, 1065-1077.

15. Engkvist, I-L, Hagberg, M., and Wigaeus-Hjelm, E. (1995). Interview protocols and ergonomics checklist for analyzing overexertion back accidents among nursing personnel. *Applied Ergonomics*, **26.3,** 213-220.
16. Wiegmann, D. A. and Shapell, S. A. (1997). Human factors analysis of postaccident data: applying theoretical taxonomies of human error. *International Journal of Aviation Psychology*, **7**, 67-81.
17. Fiske, S. T. and Taylor, S. E. (1984) *Social Cognition*. Reading, MA: Addison-Wesley.
18. Reason, J. (1997). Approaches to controlling maintenance error. In *Proceedings of the FAA/AAM 11th Meeting on Human Factors Issues in Aviation Maintenance and Inspection, San Diego, CA*
19. O'Hare, D. (2000). The "Wheel of Misfortune": a taxonomic approach to human factors in accident investigation and analysis in aviation and other complex systems. *Ergonomics, 2000,* **43**(**12**), 2001-2019.
20. Helmreich, R. (1990). Human factors aspects of the Air Ontario crash at Dryden, Ontario, in *Final Report of the Commission of Inquiry into the Air Ontario Crash at Dryden, Ontario (Ottowa: Minister of Supply and Services).*
21. Rasmussen, J. (1982). Human errors: a taxonomy for describing human malfunctions in industrial installations. *Journal of Occupational Accidents* **4,** 311-333.
22. Johnson, C. W. (2000). Proving properties of accidents. *Reliability Engineering & System Safety***, 67,** 175-191.
23. Lewko, J. H. (1998). Human factors and accident investigation: the trend for the 21st Century. *Proceedings of the 30th Annual Conference of the Human Factors Association of Canada – 1998*, Mississauga, Ontario, 355-359.
24. Lehane, P. and Stubbs, D. (2001). The perceptions of managers and accident subjects in the service industries towards slip and trip accidents. *Applied Ergonomics*, **32**, 119-126.
25. Chapman, P. and Underwood, G. (2000). Forgetting near-accidents: the roles of severity, culpability and experience in the poor recall of dangerous driving situations. *Applied Cognitive Psychology*, 14, 31-44.
26. Strauch, B. (1999). Managing aviation human factors aircraft accident investigations: lessons learned. *Proceedings of the 43rd Annual Meeting of Human Factors and Ergonomics Society*. 957-961.
27. Haslam, R. A. and Bentley, T. A. (1999). Follow-up investigations of slip, trip and fall accidents among postal delivery workers. *Safety Science*, **32**, 33-47.

28. Soares, C. G., Teixeira, A. P. and Antao, P. (2000). Accounting for human factors in the analysis of maritime accidents. In Cottam, Harvey, Pape and Talt (eds), *Foresight and Precaution.* Rotterdam, Balkema, 521-528.
29. Green, M. M., Morisseau, D., Seim, L. A. and Skriver, J. (2000). Development of an incident investigation process. *Proceedings of the IEA 2000/HFES 2000 Congress*, 4-388 – 4-390.
30. Drury, C. G., Wenner, C. L. and Murthy, M. (1997). A proactive error reporting system. *FAA/AAM 11th Meeting on Human Factors Issues in Aviation Maintenance and Inspection,* SanDiego, CA, 91-104.
31. Marx, D. A. (1998). *Learning from our Mistakes: A Review of Maintenance Error Investigation and Analysis Systems* (with recommendations to the FAA), Atlanta, Galaxy Scientific Corporation.
32. Schmidt, J., Schmorrow, D. and Hardee, M. (1998). A preliminary human factors analysis of Naval Aviation maintenance related mishaps (983111). *Proceedings of the Airframe/Engine Maintenance & Repair Conference*. Long Beach, CA.
33. Shapell, S. A. and Wiegmann, D. A. (2000). *The Human Factors Analysis and Classification System—HFACS.* Report DOT/FAA/AM-00/7, Springfield, National Technical Information Service.
34. Woodcock, K. (1999). Foolproofing: how real-world causal reasoning works against ergonomics. *Proceedings of the Association of Canadian Ergonomics*, Hull QC, 13-16 October 1999.
35. Boeing Commercial Aircraft Division (1997). *MEDA User's Guide*.
36. Rankin, W. L. (2000). The maintenance error decision aid (MEDA) process. *Proceedings of the IEA 2000/HFES 2000 Congress*, 3-795 – 3-798.
37. Drury, C. G., Wenner, C. and Kritkausky, K. (1999*). Development of process to improve work documentation of repair stations.* FAA/ Office of Aviation Medicine (AAM-240). National Technical Information Service, Springfield, VA.
38. Marx, D. and Watson, J. (2001). *Maintenance Error Causation*. FAA, Office of Aviation Medicine.
39. Bureau of Labor Statistics (1991). Washington, D.C.
40. Veteran's Administration National Center for Patient Safety web site, http://www.patientsafety.gov
41. Wenner, C. and Drury, C. G. (2000). Analyzing human error in aircraft ground damage incidents. *International Journal of Industrial Ergonomics*, **26(2),** 177-199.

# 8.0 ACRONYMS

| | |
|---|---|
| ANCOVA | Analyses of Covariance |
| ANOVA | Analysis of Variance |
| AMT | Aviation Maintenance Technician |
| BASI | Bureau of Air Safety Investigation |
| BLS | Bureau of Labor Statistics |
| CAE | Conclusion, Analysis and Evidence |
| FAA | Federal Aviation Administration |
| GLM | General Linear Models |
| HFACS | Human Factors Analysis of Postaccident Data |
| IAM | International Association of Machinists |
| MEDA | Maintenance Error Decision Aid |
| MRM | Maintenance Resource Management |
| NRR | Non-routine Repair |
| NTSB | National Transportation Safety Board |
| PERS | Proactive Error Reporting System |
| QA | Quality Assurance |
| RTS | Return to Service |
| SHELL | Software / Hardware / Environment / Liveware / Liveware system |
| SUNY | State University of New York |
| TOMES | Task / Operator / Machine / Environment / Social system |