HUMAN FACTORS EFFECTS IN THE FAA EDDY-CURRENT INSPECTION RELIABILITY EXPERIMENT

Donald L. Schurman, PhD CPE
Science Applications International Corp.
Floyd W. Spencer, PhD
FAA Aging Aircraft NDI Validation Center
Sandia National Laboratories

Introduction

High-frequency eddy current inspections are an integral part of routine maintenance checks and of directed checks for surface fatigue in aircraft fuselage skins. To investigate the reliability being achieved in high-frequency inspections in airplane maintenance facilities, an experiment using simulated lap splice joints was designed at the FAA's Aging Aircraft NDI Validation Center. The design and implementation of the experiment was sponsored by the Federal Aviation Administration Technical Center in Atlantic City, New Jersey.

The goal of the experiment was to quantify the reliability of inspections as they are routinely performed in aircraft maintenance facilities. The participants were asked to perform the inspections using their own equipment and procedures. The experiment was generally located in work areas where this type of inspection would occur, thereby subjecting the inspectors to the same environment, distractions, etc.

The experiment was taken to nine (9) maintenance facilities from March 1993 to August 1993. At each of the facilities, four (4) inspectors (or inspector teams) inspected 924 rivet sites contained in the simulation. One of the inspectors (or one of the teams) repeated the inspection, thereby permitting for repeat inspection variation to be characterized. All inspections were observed and detection data gathered by a team of monitors. One of the monitors was a human factors specialist and the other was an NDI (nondestructive inspection) specialist.

The 924 rivet sites contained 184 cracks of known and well characterized length. All of the cracks were fabricated through load cycle fatigue in aircraft skin material. Details of the fabrication are given in Reference [1]. The cracks varied in length from 14 mils to close to one inch. Most of the cracks were in the neighborhood of 100 mils. This is the size of crack specified in Boeing procedures to set-up eddy current inspection equipment.

Subjective ratings on a three-point scale were obtained from the inspectors. This was done so a relative operating characteristic (ROC) analysis could be performed to supplement probability of detection (POD) curve fitting. Here, we restrict attention to the data that include all calls made by the inspector, regardless of the subjective rating. For most inspectors, even the most lax criteria level did not include an excessive amount of false calls.

Some of the flawed rivet sites had cracks from the right side, some from the left side, and some had cracks from both sides. In marking the rivet sites containing cracks, we asked the inspectors to indicate where the crack was with respect to the rivet head. It became clear that many of the inspectors did not have the ability or the experience in pin-pointing and marking crack locations. The most common result was that only one of the cracks would be indicated at those rivet sites containing two cracks. We decided that it unduly penalized the inspector to consider the unmarked crack at a doubly cracked rivet as having been missed. In the data presented, a cracked rivet is characterized by the length of the longest crack. The effect of having two cracks is studied as a factor in the regression analysis presented in Reference [2].

Experimental and Observed Factors

Various controlled factors were incorporated into the experimental design. These factors include: off-angle flaws, painted versus bare inspection surfaces, accessibility or ease associated with the inspection task, and specimen type. Details of the experimental design, including a discussion of these and other factors are given in Reference [1].

In addition to the controlled data factors that were included in the experimental design, there are a number of identified factors that were uncontrolled, but for which information was gathered. Factors included in this case were those that, in the past, have been shown to have effects on general inspector performance. For convenience in talking about these factors we have divided them into two broad categories: Environmental and Personnel factors. The Personnel factors category has been further subdivided into Physical and Psychological factors.

The measures that were examined in this section of the research were general measures of *performance*, rather than probability of detection. POD is important to safety, but does not tell the whole story. For example, improvement of factors that improved speed of inspections without lowering accuracy would result in lower inspection costs. Human factors in procedures that led to smaller cracks being detected at adequate POD levels could also lower inspection costs by allowing more aircraft flight cycles between inspections. These sorts of results can improve overall safety by increasing the resources that can be devoted to other safety issues.

Selection Of Observational Factors

Environmental Factors. The following environmental factors were selected for measurement or recording for their effects on inspection performance shown in other research **and** for the feasibility of measuring or consistently judging them. The selected factors were:

- Housekeeping
- Temperature
- Humidity
- Lighting
- Noise levels
- Behavioral Climate
- Tool condition and availability
- Instrument and probe type

- Instrument calibration characteristics (setting)
- Procedures used
- Scanning techniques
- Changes in environmental conditions

Personnel Factors. There are a very large number of personnel factors that have been found to be more or less relevant to inspection types of tasks in previous research. We broke them into two categories: Physical and Psychological. Physical factors are more easily measured, judged, or reported (e.g., age or gender). Psychological factors are less easily measured, judged, or reported (e.g., attitude).

It should be understood that the inspectors were regular employees of aircraft maintenance facilities. The parent organizations volunteered the time of these inspectors without any recompense. In fact, due to confidentiality requirements of the research, these facilities could not even find out how well their people did. The fact that so many organizations volunteered the time and space at their own cost is a tribute to the dedication of these organizations to safety and improvement of the entire industry. The inspectors were nominally volunteers - although we know that most of them were simply assigned to the experiment in much the same way that they would be assigned to any other job.

Physical factors that were selected to be recorded were:

- Observed general physical condition
- Postures used during test
- Age
- Gender
- Previous amount of sleep
- Time on duty
- Prior activities
- Reported physical condition
- Reported fatigue level (beginning)
- Reported fatigue level (ending)

There were more psychological factors proposed overall, since the research literature is quite full of factors that have some effect upon performance (although some of those effects are quite subtle). However, it was not feasible to give these volunteers a full battery of psychological tests, so the factors were limited to those that could be obtained through self- report or that could be easily observed. In the design, however, some of these factors were obtained from both self-report and observation and, in some cases, several measures that get at the same factor were built into the system.

The *Psychological factors* that were selected to be recorded were:

- Attentiveness
- Observed attitude
- Work patterns
- Education level
- Attitude toward job
- General experience level
- Instrument-specific training
- Recency of instrument-specific training
- Type of training
- Lap-joint experience
- Instrument-specific experience
- Perceived management attitude
- Lap-joint experience recency

- Reported attitude toward experiment
- Perceived realism of test
- Reported mental condition (e.g., irritability, efficiency, depression, mental condition)
- Reported attitude during test

Results

In presenting the results we will first discuss general observations on the controlled factors, followed by some discussion of the uncontrolled factors in the experiment. Detailed analysis of the data will be presented in Reference [2].

Controlled Factors and Inspection Data Results

Table 1 presents a summary of inspection results, where the total number of finds are given in several categories along with the false call rate. The categories presented are the cracks with lengths less than 50 mils (still under the countersunk rivet head), cracks with lengths between 50 and 100 mils, and cracks that exceed 100 mils in length. The 100 mils criterion was chosen because Boeing procedures call for setting up the inspection equipment using standards with 100 mil cracks.

Table 1 Number of Crack Detections and False Calls by Inspection within Facilities. The "R" inspections are the repeat inspections.

Inspector	<50 mils	50 to 100 mils	>100 mils	Total # False calls	Inspector	<50 mils	50 to 100 mils	>100 mils	Total # False calls
A1	0	42	97	22	F1	2	52	96	9
A1R	0	39	98	0	F2	0	24	79	2
A2	0	48	98	0	F3	0	54	97	0
A3	0	47	97	0	F4	2	53	97	3
A4	0	48	95	3	F4R	0	53	98	13
B1	1	50	98	1	G1	0	13	78	5
B2	2	56	93	1	G1R	0	13	81	4
B3	0	42	97	2	G2	0	28	98	6
B4	0	38	98	0	G3	1	5	73	10
B4R	1	48	95	0	G4	6	29	89	59
C1	0	41	98	2	H1	0	41	94	17
C2	0	49	97	1	H2	0	36	93	0
C3	1	56	96	6	H2R	0	42	96	0
C4	0	22	84	1	H3	0	28	91	3
C4R	0	26	93	0	H4	1	35	90	24
D1	0	34	97	3	J1	0	22	95	1
D2	0	35	96	6	J1R	0	38	98	11
D2R	1	41	96	3	J2	1	29	87	25
D3	2	45	98	29	J3	0	42	98	3
D4	4	45	98	84	J4	1	41	96	0
E1	0	35	97	2					
E2	1	32	93	3					
E2R	1	30	94	6					
E3	0	39	96	2					
E4	2	34	97	53					
#	22	64	98	598		22	64	98	598

Table 2 gives a summary of the inspection results by Facility. Also given in Table 2 is the condition of the surface inspection at the facility. An analysis of variance (Anova) table for the number of detects in the 50 to 100 mil range is given in **Figure 1**. The Anova table summarizes the data for testing whether the facility-to-facility variation is significant when compared to the within facility (inspection -to-inspection) variation. The facility variation is significant but an examination of Table 2 indicates that the major reason for facility-to- facility differences could be attributed to facility G.

Table 2 Number of Detections and False calls by Facility

Facility	<50 mils	50 to 100 mils	>100 mils	Total # False	surface condition
A	0	224	485	25	bare
B	4	234	481	4	paint
C	1	194	468	10	bare
D	7	200	485	125	bare
E	4	170	477	66	paint
F	4	236	467	27	bare
G	7	88	419	84	paint
H	1	182	464	44	paint
J	2	172	474	40	paint
# possible	110	320	490	2990	

Figure 1 Analysis of Variance Table for Facility Effect

Source of Variation	*Sum of Squares*	*df*	*MS*	*F*	
Facility	2512.89	8	314.11	3.51	(p=.007)
Inspections	2415.	27	89.44		
Total	4927.89	35			

Figure 1

From **Table 2** it is seen that the top five facilities, with respect to 50 to 100 mil detections, includes the four facilities where the inspections were performed on bare surfaces. A t-test comparing the overall means (for the 50 to 100 mil detections) of the inspections done on bare surfaces compared to those done on painted surfaces shows statistical significance (bare mean=42.7, painted mean=33.8, p=.005).

Uncontrolled Factors

Although the list of observed and measured factors is quite long, very few of them had demonstrable effects in this research. As is often the case when factors are not controlled, some of these factors may influence inspection results, but their effects are muted or confused by interactions with other factors. For example, factors within a facility, such as attitude and experience, could have strong effects that override other facility specific effects, such as poor lighting or the discomfort of high heat and humidity.

An additional difficulty in trying to analyze the effects of human factors on performance comes from the fact that some of the facilities use 2-person teams for eddy-current inspection. Thus, there is only one performance measure for two people. How can we tell which person's characteristics are affecting the score? The answer, of course, is that we can't. However, it did appear to the monitors that the poorest person in the team dragged the team down - rather than the best person pulling the team up (weak link phenomenon). Further research would be required to determine whether this impression is valid. If it is, the implication is that the best and most effective performance is actually obtained by inspectors working solo. The cost implications of such a finding are plain. It was clear that the two inspectors did not act as independent checks on each other in performing the inspections.

Only those factors that seemed to have real effects on performance (speed, accuracy, etc.) will be discussed here. All of the factors that were examined will be discussed more fully in the final report of this project (Reference [2]).

Instrumentation Factors. The three factors of "Instrument and Probe Type", " Instrument Calibration Characteristics", and "Procedures Used" are directly related to instruments and instrument use within the facilities. In general, these factors were consistent within a given facility.

Eddy current inspection equipment ran the gamut from very old Magnaflux ED-520s and Forester 2.8s to New Zetec MIZ-22s and Nortec 19s (and one Rohmann B1 with a rotating probe). The monitors acquired the impression that the older instruments were less sensitive for detecting cracks than the newer styles of impedance plane instruments (See Hagemaier [3]).

There were several different kinds of standards blocks used for setting up the inspection equipment. Blocks specific to the lap splice being inspected and called out in Boeing procedures were the most common. However, two facilities used "universal" eddy current standards (different at each) and setup the inspection equipment in ways other than specified in Boeing NDT procedures. One facility was near the top in overall performance, while the other was the worst in overall performance.

The monitors also observed variations in *interpretation* of calibration procedures. For example, many inspectors calibrated impedance-plane instruments to have a very tall, narrow loop or a very short, flat loop - rather than the symmetrical loop called for in the procedures. Similarly, some inspectors calibrated needle-deflection instruments to small deflections for lift-off and large deflections for cracks (typically full-scale) - rather than deflection in one direction for lift-off and in the opposite direction for cracks.

Inspectors who did not know how to set up optimal impedance plane settings (90% separation of x and y axes) or optimal meter balance (liftoff in one direction and crack indications in the other) handicapped themselves by making the interpretation of readings harder.

Lighting. The inspectors were able to overcome the lack of good lighting, in many cases by using flashlights to see the rivet heads. Although lighting is not nearly so important when inspecting with EC instruments as when visually inspecting, lighting seems to affect the speed of inspection. Simply put, the cost of better lighting (portable units, etc.) could be quickly repaid in faster inspections of the same accuracy.

Personnel Physical Factors.

Although physical factors have been shown to be important in manufacturing and visual inspection, they did not seem to have large effects in this research. Quality, rather than speed, of decision-making seems most important in the use of inspection instruments. Of course, fatigue slows all functioning so that performance suffered for the few inspectors that we had who were working past 8-hours in that day.

Age. There was some consistency of age within facilities. That is, a given facility tended to have all, or mostly all, younger inspectors or all older inspectors. Thus, the possible effect of age becomes hard to separate from other facility specific characteristics. Although not statistically significant, indications are that the younger inspectors tended to perform better overall than older inspectors. The impression of the monitors was that some differences were attributable to improved training methods that have come into use in the last few years, benefiting younger, but not older, inspectors. In addition, as was once told to the first author by an old Chief Warrant Officer in a military maintenance setting, "Lots of these old-timers just know a lot of things that ain't so."

In general, the solution to older inspectors "knowing a lot that ain't so" is not so simple as just requiring refresher training. Methods must be found to encourage old-timers to adopt more effective methods and newer equipment. In other words, there is a difficult problem of getting older inspectors to *accept* rather than just *undergo* fresh training.

Previous amount of sleep/fatigue. One inspector had worked an eight-hour shift before starting the experiment, several others told us of personal and/or family problems that had kept them from good nights of sleep before coming to the experiment. Most of the midnight shift inspectors reported being somewhat tired. These inspectors probably did not work to their own potential, but the differences among inspectors were so great that it is hard to tell how much better they would have done if they had had more sleep.

Psychological Factors

Essentially all of the inspectors were attentive to the job, had reasonably good attitudes toward their work and had good work patterns. That is, all the inspectors that we observed using eddy-current equipment were there to do a good job. For this reason, many of the psychological factors did not have importance because all inspectors rated "good" on the scales.

"Behavioral Climate" and "Perceived Management Attitude" did vary quite a bit between facilities, but we found that the inspectors overcame a lot of those difficulties through personal and professional pride in their work. Factors of housekeeping, resource allocation, and morale - that have been seen to affect performance among less skilled and dedicated workers - seemed to be overcome by the inspectors that we observed.

Work Patterns. All inspectors worked diligently at the experimental task. Many of the inspectors stated that they would have taken more breaks and rest periods on a regular inspection job that was not a test. In fact, some of the inspectors who took more breaks tended to be more accurate than those who worked straight through without looking up. This finding is consistent with prior research on inspection tasks. Inspection work requires many short breaks to maintain performance.

Type of Training. Other factors being equal, the school-trained inspector participating in the experiment was most accurate and performed the inspection in a timely manner. Inspectors who had taken good quality local classroom training (typically ASNT Level III instructed) on their instruments also did well. On-Job Training, even formalized on-job training, did not seem to produce the quality of inspectors that formal classroom training produced. It must be noted, however, that training in the theory of eddy-current instrumentation was not nearly so useful as training on the proper calibration and use of a specific instrument.

Specific Experience. Overall experience in NDT inspection appeared to be less important than experience with the particular instrument being used. In some cases, the inspectors had been well-trained on other instruments and spent the time and effort to familiarize themselves with the available instrument(s). In other cases, the inspectors were not familiar with new or upgraded instruments, which seemed to hamper their performance.

Summary and Conclusions

The goal of the Eddy Current Inspection Reliability Experiment was to quantify how reliably inspections can be carried out in aircraft maintenance facilities. Factors that might affect the reliability and that could be controlled were designed into the experiment. One of these factors, inspection surface condition, was shown to be significant using categorical detection data. The full quantification of the design factor effects will appear in a final report [2].

Data were also collected on various uncontrolled factors related to the inspectors and the inspection environment. Variations in many of these factors were observed and have been briefly described. Facility correlations and the particular mix of the observed factors, does not allow for statistical significance to be demonstrated for the effect of the uncontrolled factors. However, suggested influences of some of the factors are consistent with published research and have been briefly discussed.

The facilities who supported this research and the inspectors who participated were helpful in every way and seemed to want the program to succeed. The inspectors felt that the mock- up of lap joint inspections were a fair representation of the job and that, if anything, they would do a little better on the mock-ups than on the real thing. The mock-up was more uniform and regular than aircraft that have seen a lot of cycles of service. The inspectors were obviously doing their best to perform well on what they considered a "test". Nonetheless, we feel confident that this research reflects the general level of inspection quality that is found in the American Air Carrier Industry today.

References

Spencer, Borgonovi, Roach, Schurman, and Smith, "Reliability Assessment at Airline Inspection Facilities, Volume II : Protocol for an Eddy Current Inspection Reliability Experiment," May 1993, DOT/FAA/CT-92-12, II.

Spencer, F. W. and Schurman, D. L., "Reliability Assessment at Airline Inspection Facilities, Volume III : Results of an Eddy Current Inspection Reliability Experiment," to be published as DOT/FAA/CT-92-12, III.

Hagemaier, D.,"Reliable Supplemental Inspections of Aging Commercial Aircraft," 1987 ASNT Fall Conference, Atlanta Georgia, October 5-9, 1987.