VISUAL INSPECTION RELIABILITY: WHAT WE KNOW AND WHY WE NEED TO KNOW IT

Colin G. Drury, Ph.D.
University at Buffalo
Professor of Industrial Engineering

Abstract

Inspection reliability is the key to continuing airworthiness of the civil aviation fleet. First we must have quantitative knowledge of inspection reliability to determine inspection intervals, then we must find ways to improve this reliability if we are to keep costs under control while ensuring safety. This paper shows where we get our data and models for inspection reliability, concentrating on the human inspector within the inspection system. Three examples, from FPI, borescope inspection and visual inspection, show how such data and knowledge can be applied to derive Good Practices that have since been adopted in the industry. But our knowledge of inspection reliability is incomplete, so that continuing studies are required to support the continuous improvement demanded by quality systems. Two examples of such studies show how off-line experiments can be combined with field observations to further our understanding of reliability, and can lead to implementable conclusions.

1. Why Do We Need to Know About Inspection reliability?

Airworthiness of civil aircraft depends upon a process by which a team composed of aircraft manufacturers, regulators and one or more airlines predict possible system failures. This process, Maintenance Steering Group 3 or MSG-3, considers possible failure pathways (for example, in structures, engines, avionics) and for each pathway determines a recovery strategy. For structural failure in airframe or engine this is often regular inspection to assure detection. The concern here is with the reliability of the primary failure recovery system for aircraft and engine structural inspection: regular inspection to assure detection.

Inspection systems are designed to detect all structural in a timely manner, i.e. before the failure has a catastrophic effect on structural integrity. For example, crack growth rates in engine and airframe critical components can be predicted probabilitistically from material properties and applied stresses, so that the MSG-3 process can schedule inspections before a potential crack becomes dangerous. However, the detection system has certain limits on size crack that can be detected, so that MSG-3 typically schedules several inspections between the time the crack becomes detectable and the time it becomes dangerous. If too many inspections are scheduled, the costs are driven up in a highly-competitive industry, and the risk of collateral damage is increased due to the handling activities involved in the inspection process itself. Conversely, if too few inspections are scheduled, the probabilistic rate of the crack growth prediction process may combine with the probabilistic nature of the detection process to cause dangerous cracks to remain undetected. Spectacular failures of this inspection process have occurred both for aircraft structures (Aloha incident, Hawaii 1988) and engine components (Pensacola incident, Florida 1997).

The MSG-3 process thus requires quantitative data on inspection reliability to function correctly. In addition, no rule-based prediction system can foresee all possible malfunctions, so that once an aircraft is in service, regular detailed inspections are made of the whole structure to discover any unexpected cracks. When such "new" cracks are found, the information is typically shared between manufacturers, operators and regulators in the form of supplementary inspections. Similar considerations apply to other failure modes such as corrosion.

This whole reliability assurance process thus rests upon an inspection system which checks both points where malfunctions are expected and points where they are not expected, for a variety of malfunctions. For good reasons, human inspectors are part of this inspection system, so that human inspection reliability is an essential element in ensuring structural integrity, and hence airworthiness. There is minimal useful information in descriptive studies: to predict performance and hence airworthiness we need quantitative reliability data.

2. What Do We Know So Far?

The inspection task itself is classified in aviation as either Visual Inspection or non-destructive inspection. Regulatory bodies have issued formal descriptions of both of these tasks (e.g. Bobo (1989)[a] for the FAA), and both have somewhat different characteristics in aviation

Non-destructive inspection (NDI) comprises a set of techniques to enhance the ability to detect small and/or hidden malfunctions. One set of NDI techniques are those which enhance what is essentially still a visual inspection task, for example X-ray, fluorescent particle, magnetic particle or D-sight. They show cracks that are very small (fluorescent particle) or hidden within other structures (X-ray). Apart from the steps necessary to ensure a good image, they have many of the human interface characteristics of visual inspection. Visual inspection is much more common, comprising 80% of all inspection[b]. It consists of using the inspector's eyes, often aided by magnifying lenses and supplementary lighting, as the detection device. Inspectors must visually scan the whole structure of interest, typically using portable mirrors to examine areas not directly visible. Whether the task is categorized as Visual Inspection or NDI, its aim is to detect flaws (indications) before they become hazardous. There are three sources of knowledge about inspection reliability, and in particular human reliability, in inspection.

2.1 Field Studies

We have measurements of inspection reliability from a number of field studies of aircraft inspection tasks over the years. Perhaps the earliest was by Lock and Strutt (1985),[c] which performed detailed Task Descriptions of sample inspection tasks and used these in Human Reliability models to understand the areas where inspection was potentially vulnerable to human error. A more extensive sample of inspection tasks by Drury, Prabhu and Gramopadhye (1990)[d] used a Task Analysis format to look for areas where task demands potentially exceeded human capabilities, i.e. where errors are most likely. This study in fact structured much ensuing work by University at Buffalo and others on improving inspection reliability, by pointing out where relatively small changes could have large effects. These included design of instructions, lighting for inspection, training / retraining of inspectors and work in restricted spaces. As part of this work we have developed better training programs (e.g. the ASSIST program[e]), a much-used job aid for writing documentation with less susceptibility to misunderstanding,[f] and a guide to choice of lighting systems.[g]

2.2. Off-line Studies

These are controlled studies in an off-line environment, which may include work using realistic conditions on a real aircraft in a hangar, but where the inspection is not part of the regular inspection of the aircraft for determining airworthiness. Good examples are the studies over several years at Sandia National Laboratories AANC. The first, the ECRIRE study,[h] followed directly from the Aloha incident in testing inspectors using their own eddy current equipment to detect cracks in panels specially manufactured to conform to a typical fuselage skin rivet joint. There were large differences between inspectors (as usually found), no consistent effects of one-person versus two person teams and some differences between testing organizations. In addition, working at a comfortable height gave better detection performance than working closed to floor level.

Other studies in the same series have been part of the Visual Inspection Research Program, VIRP. The first used 12 experienced inspectors to perform six inspection tasks on a B-737 test bed fuselage. Again, individual differences were large, but the study did give practical measures of the probability of detecting cracks and corrosion on a full-size airframe. Such data is exactly what is needed for ensuring continuing airworthiness.

2.3 Analog Studies

These comprise field or off-line studies from domains other than aircraft airworthiness inspection. Human factors analyses of inspection tasks have been published since the 1950's and 1960's with a steady evolution of approaches. Early studies (e.g. Thomas and Seaborne, 1961[i]) tended to be rich and holistic descriptions of inspection tasks. These showed for example that inspectors organize their perspectives so as to enhance subtle task relevant visual or auditory cues and suppress what a novice would perceive as salient cues.

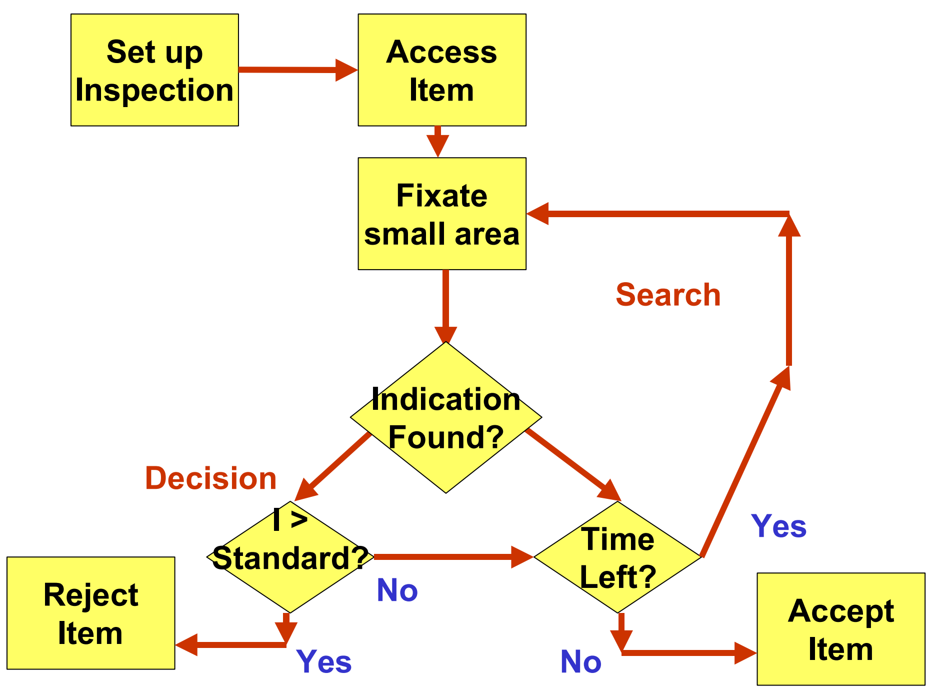
The next wave of work measured human performance in a variety of inspection tasks, typically in terms of the two possible errors: missed defects and false alarms. Reviews of this work are readily available.[j,k] Table 1 classifies some of the factors found to affect inspection performance. Following such studies, and indeed overlapping them, were model-oriented studies treating inspection as either a signal detection task[l,m] or a visual search task.[n,o] The advantage of such approaches is that they can use the underlying models to predict which variables are most and least likely to affect inspection performance. They also allow succinct descriptions of tasks and task performance, potentially leading to quantitative models. For example, studies of aircraft inspection often provide Relative Operating Characteristic (ROC) curves relating miss rate to false alarm rate for a given defect type.

An inspection model combining search and decision[p] (Drury, 1975) can also be helpful in understanding the inspector's tasks in inspection. This model, summarized in Figure 1, shows an inspector searching an item by repeated fixations of small areas. If an indication (potential defect) is found, a decision task takes place to determine whether the indication should be classed as a reject. If not, or if the fixation found no indications, search continues. The inspection task stops (or moves to the next item) when there is no further time left for inspection, usually because of the inspector's stopping policy. This model allows us to specify the variables affecting each stage. Thus, peripheral visual acuity should affect fixation area and thus, search performance.[q] Conversely, the decision stage should be affected by cost and probabilities of the decision outcome.[r] Overall, this model has been useful in interpreting the speed/ accuracy tradeoff in inspection.[s]

Table 1. Summary of inspection findings using Task / Operator / Machine / Environment / Social Model

TOMES Component	Typical Findings from Inspection Studies
Task e.g. procedures, instructions, workcards, feedback	Instructions given to the inspector have a great effect on both p (detect) and p (false alarm). In addition, feedback information to the inspector has large positive influences on performance.[t]
Operator e.g. individual inspector characteristics	Some general characteristics of "good" inspectors, such as field independence and peripheral visual acuity. Often each inspection task shows performance correlations with different individual characteristics.
Machine e.g. job aids, enhanced vision systems, magnifiers	Equipment such as semi-automated visual inspection systems improve performance when well-integrated with human functions.[u] Enhanced vision systems, such as magnification or lighting aids sometimes help, sometimes do not. Providing visible comparison standards improves decision.
Environment e.g. lighting, thermal, noise	Some effects, but only at relatively extreme values and with long exposure times.
Social e.g. interactions with other people in system	Job design is important. Inspectors tend to feel their jobs isolate them from others. Expectations of others can have large effects on what gets reported as fault.

Figure 1. Search and Decision Processes in Inspection



A flavor of the findings of this tradition can best be given through a simple model of human factors each element of which represents a key component of the human/ machine system. How each component interfaces with the individual considered determines the sources of both inspection reliability and human errors. Table 1 summarizes industrial inspection findings using this model of human factors.

3. Where Are We Now?

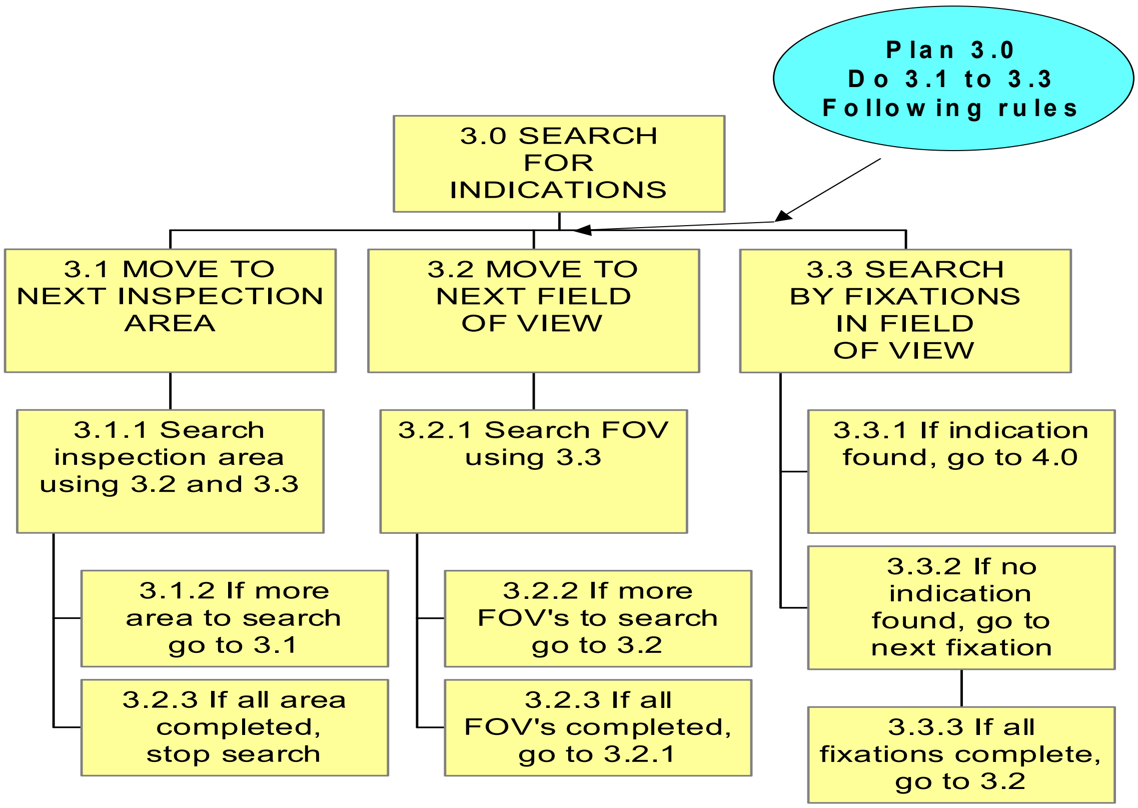
The data and models from the three sources above can be brought together to produce state-of-the-art applications to different aircraft inspection tasks. Three such studies have been completed so far, on Fluorescent Penetrant Inspection (FPI), on Borescope inspection and on Visual Inspection. The series of application studies were prompted by the engine component failure noted earlier (Pensacola incident, Florida 1997). One proximal cause of that failure of a JT8-D hub was that an existing crack was remained undetected during a routine FPI procedure. The crack was hidden in a long bore, but should in fact have been visible using the FPI procedure.

Each of these studies used a common format, with a detailed Task Analysis to find potential mismatches between task demands inspector capabilities that could lead to errors. Figure 2 shows the top level of this Task Analysis for borescope inspection, from Drury and Watson (2001).[v] The Task Analysis was used with detailed job observation and interview to produce a comprehensive set of Good Practices for that task, e.g. Table 2. There were 88 such Good Practices for FPI, 56 for borescope inspection and 61 for visual inspection. Most importantly, our increasing knowledge of inspection reliability allowed the incorporation of a reason for each Good Practice (the **Why** column in Table 2 below). Thus users can go beyond a prescription of what they should do to become more knowledgeable about the technical background to the recommendations.

These Good Practices can be used as part of a self-study checklist for inspection organizations, e.g for FPI reading of parts to allow users to apply the findings across many tasks easily:

Check	Yes	No	Comments
• Can the part be re-positioned easily to bring eyes to the correct position to inspect?	☐	☐	________________
• When re-positioning, can the inspector manipulate the carrier, part, light(s), swabs, solvents, loupe together as needed?	☐	☐	________________
• When re-positioning, can the inspector manipulate the carrier without it swinging?	☐	☐	________________

Figure 2. Example of a Task Analysis for one Task of Visual Inspection



Note that the full text reports can be found on **http://hfskyway.faa.gov** and that the site also allows access to all previous reports on inspection reliability produced under the FAA's Office of Aviation Medicine initiative since 1989.

4. What Do We Need To Know Next?

The three Good Practices reports have applied a wealth of knowledge from the three sources given in the second section above, and the first of these is now being actively used in several organizations, including an airline, various engine manufacturers and the armed forces to improve their FPI processes. However, these report also provided pointers to areas where we still lack information and models to help us make concrete recommendations. Each report, in fact, had to include a section on Research Needs, which tells something of the complexity of inspection reliability after over a decade of quite intensive effort.

Table 2. Selection of Good Practices for Borescope Inspection

Process	Good Practice	Why?
4. Search	Provide memory aids for the set of defects being searched for.	1. Search performance deteriorates as the number of different indication types searched for is increased. Inspectors need a simple visual reminder of the possible defect types. A single-page laminated sheet can provide a one-page visual summary of defect types, readily available to inspectors whenever they take a break from the borescope task.
4. Search	Provide training on the range of defects possible, their expected locations and expected probabilities to guide search.	1. If inspectors know what defects to look for, how often to expect each defect, and where defects are likely to be located, they will have increased probability of detection. 2. If inspector rely on these feed-forward data, they will miss defects of unexpected types, in unexpected locations, or unusual defects. Training and documentation should emphasize both the expected outcome of inspection and the potential existence of unusual conditions.
4. Search	When an indication is found, or the inspector is interrupted, ensure that inspector can return to exact point where search stopped.	1. Loss of situation awareness during blade rotation and after interruptions can lead to missed blades or missed areas on a blade. With visual inspection it is possible to mark the current point in the search, e.g. with a pen or attached marker. For borescope inspection this is not possible, but a means of locking the system when an interruption occurs will lead the inspector back to at least the current FOV.
5. Decision	Ensure that inspector's experience with all defect types is broad enough to recognize them when they do not exactly match the prototypes illustrated	1. In recognition of a defect, inspectors use their experience and any guidance from the documentation. Illustrations show typical versions of a defect that may be different in appearance from the indication seen on the engine. Inspectors' experience should allow them to generalize reliably to any valid example of that defect type. In this way, defects will be correctly recognized and classified so that the correct standards are used for a decision. 2. Training programs need to assist the inspector in gaining such wide-ranging examples of each defect type. They should use multiple, realistic indications of each defect type to ensure reliable recognition.
5. Decision	Use consistent names for all defect types	1. Unless indications are correctly classified, the wrong standards can be applied. This can cause true defects not to be reported, and false alarms to disrupt operations unnecessarily.

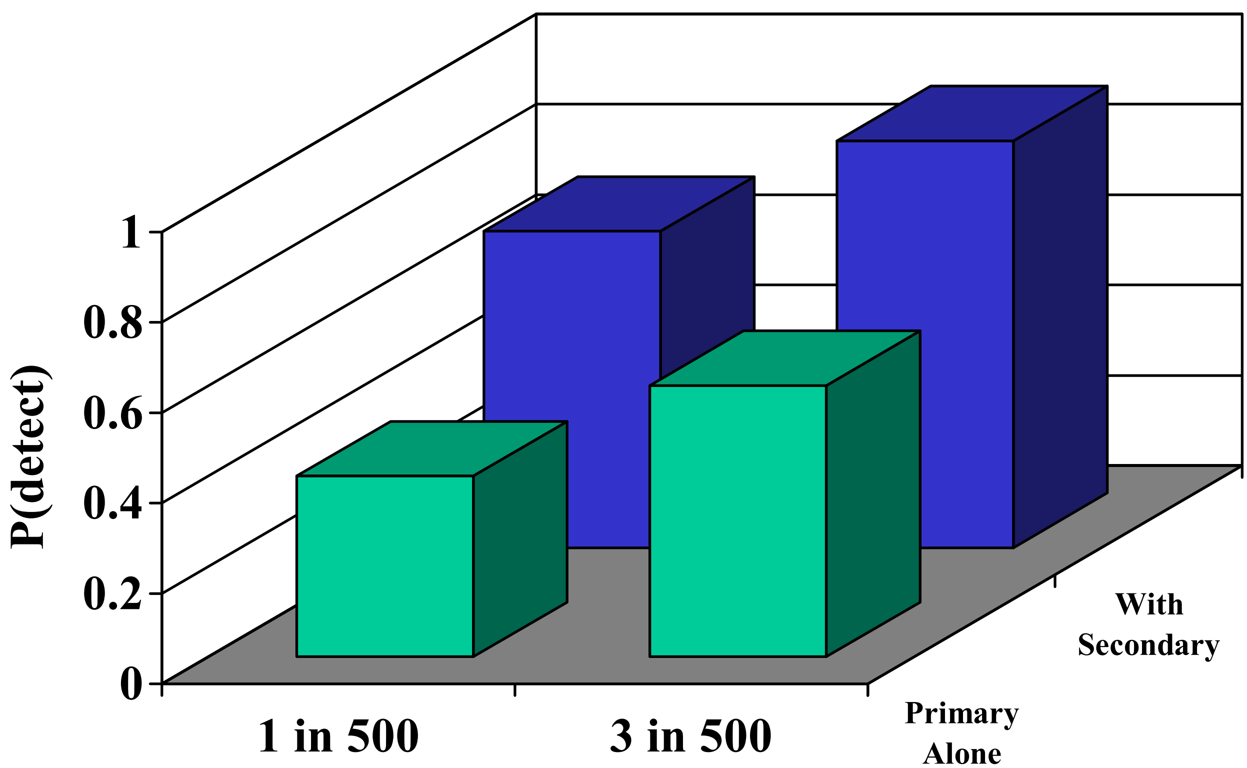
Future data collection needs are being addressed through all of the sources in Section 2 of this paper. Two examples follow where the needs have been addressed, at least initially, through experimental work.

4.1 How can we improve detection of very rare defects?

The FPI report noted that for critical rotating components on engines, the probability of actually finding a defect was extremely low. Many inspectors will never see one of these defects in their working lifetime, a problem repeated in many inspection tasks beyond FPI. We know from models of human inspection performance that inspectors tend to behave in an "optimal" manner, reducing their reporting of a defect as the defect becomes rarer. This increases the chance of being correct, but it means that as overall quality improves, the detection of defects will suffer. There is much data (e.g. Wickens, 1992)[w] showing this effect when the defect rate is 40% or 20% or 5% but not at the levels of defect rate seen in many high-quality inspection tasks where defect rates are far below 1%. In 2001 we ran a study[x] to measure this effect at the low defect rate of 1 in 500. We tested participants in a laboratory analog where they had to search computer screens for targets, and only a single target appeared on the 9th of 10 experimental sessions. Using 40 participants we tested the effect of adding more real targets (3 in 500) and of adding a more easily detectable but different target.

The results showed that both adding more real targets or adding a secondary target improved the probability of detection, and the significant part of this improvement came from changing the number of real targets, as shown in Figure 3. Clearly even a short off-line experiment allowed us to make some recommendations. The defect rate does indeed matter, as expected, so that any automation that helps remove known defect-free items from the inspection system will help p(detect). Also, having the inspector search for an unrelated defect does not distract but actually improve inspection reliability.

Figure 3. Increase of p(detect) with higher defect rate and addition of secondary target



4.2 How can we improve reliability of highly practiced inspection tasks?

As one of our earlier field studies of visual inspection discovered,[y] when an inspector performs the same complex task repetitively, deviations from correct practice do occur. The task we studied was the overnight inspection of a typical passenger aircraft, a task that required the AMT to check 108 items on the interior and exterior of the aircraft. Pearl and Drury found that inspectors tended to work the task from memory rather than from the official paperwork, and that they performed the 108 tasks in an order that made sense to them, but was not the order specified in the paperwork. Can we change the task to improve the reliability?

This we tested[z] by setting up all of the 108 tasks as charts around a room, so that we could use engineering students in place of AMTs. All the participants had to do was access each chart and check whether the function shown was correct or had a defect, thus eliminating the need for the long training undertaken by AMTs on the real task. Each of the 24 participants was trained to perform the task correctly using paperwork supplied. They were then brought back each day for 10 days to repeat the task. We measured whether or not the followed the procedure correctly, and how many of the defects they missed. Based on our knowledge of human functioning, we tested two conditions, one with a Functional sequence and one with a Spatial sequence. In the Functional condition, all of each type of check were grouped together, as on the actual documentation used by a number of airlines. Thus, all oxygen bottles were checked, then all life vests, etc. In contrast, the Spatial layout had the participant check all items in the cockpit, then in the front cabin, then the aft cabin etc.

There were very few missed errors, only 2 (functional) and 4 (spatial) out of the 72 possible occasions, so layout did not affect overall reliability. But there was a large effect of sequence errors, where participants performed steps out of the specified order. For the Functional layout there were 1068 such errors, while for the Spatial layout, this was reduced to 332, almost a 70% reduction in errors.

Thus, while in this experiment people were very reliable, finding 96% of the errors, their behavior mirrored that of the AMTs in performing many steps out of sequence despite instructions to follow the paperwork. The practical finding is that reorganizing the documentation to reflect the way people naturally perform such task using a spatial sequence can improve the reliability of compliance dramatically.

5. Where Does This Leave Us?

Continuing airworthiness still depends upon the finding of defects in a timely manner before they grow to dangerous sizes. To design a system that ensures airworthiness, reliability of inspection must be first predicted (to specify the appropriate inspection intervals) and then improved. We have given an overview of what is known about inspection, and shown procedures for applying this knowledge to give Good Practices. Two examples have demonstrated how we are pursuing more complete and quantitative knowledge about inspection reliability by merging results from field studies, off-line experiments and analog studies in other domains. The quest for improved knowledge of inspection reliability continues, but there is much that is known now and can be put to immediate use.

References

a. Bobo, S., 1989, Communication and transfer of non-destructive inspection information. In *Human Factors Issues in Aircraft Maintenance and Inspection - Information Exchange and Communications*, Second Federal Aviation Administration Meeting. Washington D.C.

b. Goranson, U. F. and Rogers, J. T., 1983, Elements of Damage Tolerance Verification. *12th Symposium of International Commercial Aeronautical Fatigue*. Toulouse.

c. Lock, M. W. B. and Strutt, J. E., 1985, Reliability in in-service inspection of transport aircraft structures, CAA Report 95013, London, Civil Aviation Authority.

d. Drury, C. G., Prabhu, P. and Gramopadhye, A., 1990, Task Analysis of Aircraft Inspection Activities: Methods and Findings, *Proceedings of the Human Factors Society 34th Annual Conference,* pp. 1181-1185.

e. Chen, S., Gramopadhye, A. K. and Melloy B. J., 2000, The Use of a High-Fidelity Simulator to Improve Aircraft Inspection Performance: The ASSIST Program, *Proceedings of the Human Factors and Ergonomics Society 44th Annual Conference, V*ol*. 2,* pp. 491-496.

f. Drury, C. G. and Sarac, A., 1997, A design aid for improved documentation in aircraft maintenance. *Proceedings of the 41st Annual Human Factors and Ergonomics Society Meeting,* Albuquerque, NM, pp.1158-1162.

g. Reynolds, J. and Drury, C. G., 1993, An evaluation of the visual environment in aircraft inspection. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, Seattle, WA, pp. 34-38.

h. Spencer, F. and Schurman, D., 1995, *Reliability Assessment at Airline Inspection Facilities. Volume III: Results of an Eddy Current Inspection Reliability Experiment, DOT/FAA/CT-92/12*, Atlantic City, FAA Technical Center.

i. Thomas, L. F. and Seaborne, A. E. M., 1961, The sociotechnical context of industrial inspection, *Occupational Psychology*, pp. 35, 36-43.

j. Drury, C. G., 1992, Inspection Performance. In G. Salvendy (Ed.), Handbook of Industrial Engineering. New York, John Wiley and Sons, pp**.** 2282-2314.

k. Megaw, E. D., Alexander, C. J. and Richardson, J., 1979, Fault mix and inspection performance, *Int. J. Prod. Res***., 17(3),** pp. 181-191.

l. Harris, D. H., 1969, The nature of industrial inspection. *Human Factors,* **11**(**2**), pp. 139-148.

m. Drury, C. G. and Addison, J. L., 1973, An industrial study of the effects of feedback and fault density on inspection performance. *Ergonomics* (**l6),** pp. l59-l69.

n. Kundel, H. L., 1975, Peripheral vision, structured noise and film reader error. *Radiology,* **114,** pp. 269-273.

o. Drury, C.G., 1990, Inspection Performance, In G. Salvendy (Ed), Handbook of Industrial Engineering (2nd Edition), John Wiley & Sons, New York, pp. 2282-2314.

p. Drury, C. G., l975, Inspection of Sheet Materials - Model and Data, *Human Factors*, **l7**, pp. 257-265.

q. Courtney, A. J., 1984, A search task to assess visual lobe size. Human Factors, **26**(3), pp. 289-298.

r. Chi, C.-F. and Drury, C. G., 1998, Do people choose an optimal response criterion in an inspection task? IIE Transactions (1998), **30**, pp. 257-266.

s. Drury, C. G., 1994, The speed-accuracy trade-off in industry. Ergonomics, **37,** pp. 747-763.

t. Gramopadhye, A. K., Drury, C. G. and Sharit, J., 1997, Feedback strategies for visual search in airframe structural inspection. *Int. Journal of Industrial Ergonomics*, **19(5)**, pp. 333-344.

u. Hou, T.-S., Lin, L. and Drury, C. G., 1993, An empirical study of hybrid inspection systems and allocation of inspection function. *International Journal of Human Factors in Manufacturing*, **3**, pp. 351-367.

v. Drury, C. G. and Watson, J., 2001, *Human Factors Good Practices in Borescope Inspection*. Office of Aviation Medicine, FAA, National Technical Information Service, Springfield, VA.

w. Wickens, C. D. (1992). *Engineering Psychology and Human Performance*. NY, Harper Collins.

x. Panjwani, G. K. (2001). Effective interventions in rare event inspection. Unpublished Ph.D. Dissertation, University at Buffalo, SUNY, Buffalo, NY.

y. Pearl, A. and Drury, C. G., 1995, Improving the reliability of maintenance checklists. *Human Factors in Aviation Maintenance - Phase Four, Progress Report, DOT/FAA/AM-93/xx,* National Technical Information Service, Springfield, VA.

z. Larock, B. P., 2000, The effects of Spatial and Functional Job-Aid Checklists on a Frequently Performed Aircraft Maintenance Task, Unpublished Ph.D. Dissertation, University at Buffalo, SUNY, Buffalo, NY.