QUANTITATIVE EVALUATION OF NDE RELIABILITY

W.E. Gardner and R.A. Murgatroyd

NRL, Risley
UKAEA, Warrington, England

INTRODUCTION

A comprehensive reliability programme is being performed by the UKAEA and CEGB which is studying aspects such as inspection procedures, equipment, and data interpretation and reporting(1). The influence of management and organisational aspects, and psychological and environmental factors are also being investigated, and the importance of these aspects has recently been highlighted by Behravesh et al(2). The information produced will enable recommendations to be made on methods of eliminating or controlling potential errors. These recommendations should, when coupled with a demonstration of the capability of the procedures, lead to objective and auditable assurance of the overall reliability of the inspection.

The need for improved reliability is demonstrated by the variability that occurred in the results of international round robin inspection exercises such as PISC II and the Defect Detection Trials which have been conducted over the last few years, aimed at evaluating the capability of ultrasonic inspection techniques to detect, locate and size small flaws in test assemblies simulating closely some of the complex geometries found in PWR's and BWR's. These exercises have demonstrated the very good capability of ultrasonic techniques to inspect thick section weldments and components, and several teams have achieved excellent results.

However, in some instances errors occurred that could be attributed to equipment design, mistakes by an operator, or difficulties in interpreting complex ultrasonic sizing data. The PISC II results have, therefore, been analysed retrospectively with the objective of identifying as far as possible, the factors that contribute to the reliability of ultrasonic inspection(3). The first part of the analysis was aimed at rating performances in terms of detection capability, flaw location accuracy and sizing accuracy, and has demonstrated the variability that can occur between teams. The results provide a basis for a detailed investigation of the factors influencing performance and this is now in progress.

Some conclusions from the reliability studies are presented and the need for future investigations, focused on the key factors and issues, is outlined.

THE RELIABILITY PROGRAMME

The principal objective of the reliability programme is to provide information which will enable an effective inspection to the performed. This aim incorporates both reliability aspects and cost effectiveness and is intended to cover the reliability of all stages of preparing for and carrying out an inspection. Clearly the specification of the task, targets, standards, techniques and equipment must be adequate to ensure that the inspection system has the required inspection capability. Hardware and software reliability need to be assured. Also the use of the system, which is normally specified in inspection procedures, must ensure that the designed capability is not downgraded during the actual inspection due to human error, or inadequate procedures or training. In the programme a methodology has been developed for analysing proceduralized tasks for human error in a systematic and demonstrably auditable manner. The method, termed SHERPA (Systematic Human Error Reduction and Prediction Analysis)(4), has been applied to plant inspection procedures in the UK. The principles of SHERPA will be outlined first, then the method of application and results achieved in the case studies will be illustrated.

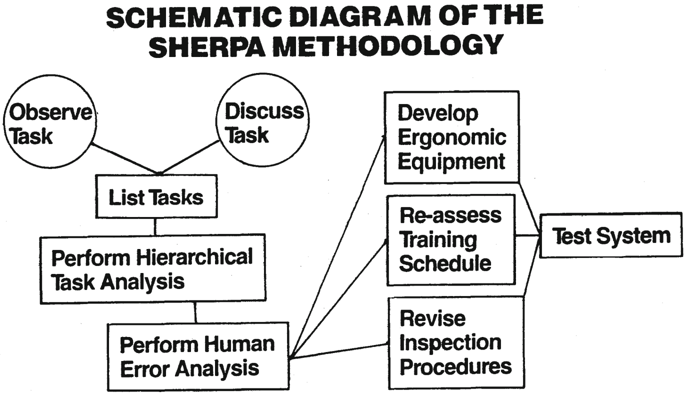

Fig. 1. Systematic Human Error Reduction and Prediction Approach (SHERPA)

The SHERPA methodology

The SHERPA methodology(4) can be applied to any system where reliability depends on Human Reliability. In applying the methodology it is essential to produce a plan of the tasks involved and the outcome is greatly improved by discussion with the operators and observation of the actual inspections. The analysis techniques involved are indicated schematically in Figure 1.

The first step is to identify the tasks necessary to achieve the system objective. A hierarchical task analysis (HTA) is then performed with the purpose of:

(i) obtaining a systematic description of the overall structure of the task in a hierarchical form

(ii) describing the actual actions to be performed and the plans which guide these actions

(iii) classifying the tasks in terms of the types of the actions required

(iv) obtaining information which can subsequently be used for the design of procedures, training and equipment in order to facilitate task performance.

The HTA is constructed by interviewing managers and operators, and by observation of the performance of the task and is then used as the basis for a Human Error Analysis (HEA) in which each operation is assessed for error sources. The objectives of this analysis are to:

(a) identify possible unrecovered error modes that could occur at each of the task steps documented by the HTA

(b) postulate error recovery mechanisms

(c) identify consequences of unrecovered error modes

(d) identify psychological mechanisms underlying these error modes

(e) develop recommendations for procedure training and equipment design to minimise the probability of occurrence of errors and maximise the likelihood of recovery.

Using the information gathered from the hierarchal task analysis and the human error analysis, SHERPA is applied to develop optimised procedures containing strategies for eliminating or recovering from error, to formulate recommendations on training requirements and to comment on ergonomic aspects of equipment design.

Evaluating SHERPA Effectiveness

The effectiveness of the SHERPA methodology has been evaluated in assessments made on specific inspections in the UK. The overall outcome of such assessments demonstrates either the capability of the analyst to apply the SHERPA methodology effectively, or the level of reliability of the procedures; or both. For instance, if only a few insignificant errors are identified this could mean that the procedures are reliable, as may be the case for a system that has evolved over several years, or that the performance of the analysis team was inadequate. If on the other hand, significant sources of error are identified then clearly the analysis has been of value. Two separate tests have been made of the performance of SHERPA. The first was on a planned inspection being performed at the UKAEA laboratories at Risley. In this the calibration and inspection procedures were analysed to detect sources of human error and to identify recovery stages. Appropriate checking mechanisms were recommended for human activities which were subject to unrecovered ie undetected, error sources. The performance of the improved procedures was tested with a new team in calibrating the inspection procedures.

The conclusion reached was that the new procedures improved reliability and could be applied effectively by the inspection teams.

A second more extensive assessment of the value of the methodology has been made on the procedures for inspecting a component in a 1890 MW(e) hydro-electric power plant in the CEGB(4). To enable the effectiveness of the analysis to be evaluated, inspection procedures for the plant dating from 1981 were used in a retrospective analysis in which the analysts were precluded from having knowledge of subsequent changes to procedures or any errors that had occurred. The results of the analysis were formally reported and then were discussed with the plant inspectors to establish which errors, if any, had occurred in practice and which were considered feasible.

The results revealed that SHERPA had predicted 99% of the human errors that were known to have occurred since 1981 or were considered by the inspectors to be feasible. About half of the data consisted of errors that had actually occurred; of these, 98% were predicted.

Many of the human errors were trivial and were recovered almost immediately. To assess the significance of the errors, they were graded according to time lost before recovery. Those causing minor losses of less than one hour represented 73% of the total. The remainder, involving the loss of an eight hour shift or more, totalled over 40 actual errors and a similar number that were considered feasible. Comparing the saving in working time with the expense involved in the analysis it is concluded that the application of SHERPA at an early stage would have been cost effective. From a reliability viewpoint SHERPA also identified a small number of significant errors that would not be recovered during the inspection stage, and the overall implications of these are being investigated.

THE PISC II RETROSPECTIVE ANALYSIS

Defect Detection

In the PISC II round robin exercise, the criterion for flaw detection is that the defect volume reported by a team must overlap the reference flaw volume. In practice, all teams reported some flaws close to the reference flaw but not overlapping it. Therefore in the retrospective analysis of PISC II results(3) an assessment was included of the effect of mis-location errors on defect detection, in which the flaw size reported by a team was increased until overlap occurred (a limit of 150mm was set on the increase). The average value of the Dimensional Increase for Overlap (DIO) from all the flaws in PISC Plate 3 was then evaluated for each team to give an indication of the location and sizing accuracy. This criterion gives the maximum flaw detection number for each team. The maximum flaw detection number is plotted in Figure 2 versus the average DIO (dimensional increase for overlap) for each team included in the groups of procedures defined as ASME 10% DAC, ASME 20% DAC and special procedures groups(1). Considerable scatter exists in the data, however, a least squares fit is shown for each group of procedures. An important observation from Figure 2 is that all three lines have a negative gradient. Thus, although a team's detection performance is improved by increasing the reported flaw size (large DIO), in many cases this does not result in a 100% detection. In effect, this implies that poor detection is coupled to poor location and sizing accuracy for flaws detected.

There is a tendency for the procedure groups to rank in the order:- Special procedures; ASME 10% DAC; ASME 20% DAC; but there is considerable scatter within a group, and for the better teams in each there is no significant difference in performance. Assuming that the detection procedures for any one group are similar, then the variability in defect detection within a group suggests a strong operator influence on the reliability of the initial data gathering and detection stage.

Further analysis is proceeding by means of structured questionnaires to the PISC participants to determine detailed information on detection and sizing procedures etc, and from the replies received it is considered that more detailed discussions are necessary.

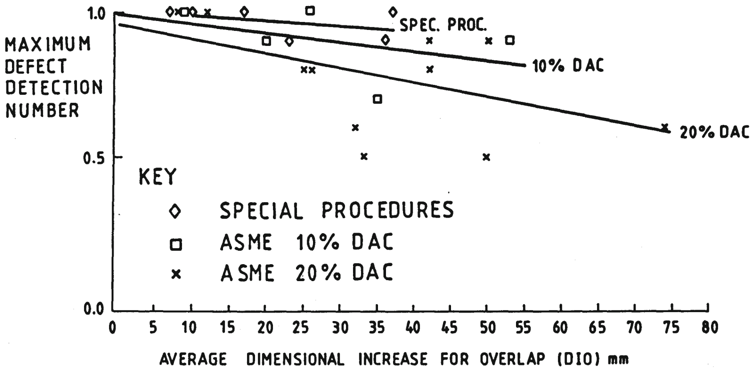

Fig. 2. Detection as a Function of Average Dio

Defect Classification and Sizing

Following the sizing of flaws by the PISC teams, each defect has been classified by the Reference Laboratory (JRC, ISPRA) as acceptable or rejectable, according to ASME rules. In this analysis of flaw classification performance(3), the parameters considered are; detection performance on rejectable flaws (DDFR); the performance of a team in correctly rejecting rejectable flaws (CRF); and the incorrect rejection of acceptable flaws (RAF). RAF is an indication of potential inefficiency in terms of unnecessary repair.

The results for CRF as a function of DDFR are show in Figure 3. Eight teams detected all the rejectable flaws, 5 from the special procedures group, 2 from 20% DAC and 1 from 10% DAC, but only three teams correctly rejected all the flaws (2 from special procedures, 1 from 20% DAC). The performance of the other teams is influenced to a greater or lesser extent by undersizing. The poor performance of ASME 50% DAC procedures is apparent in Figure 3.

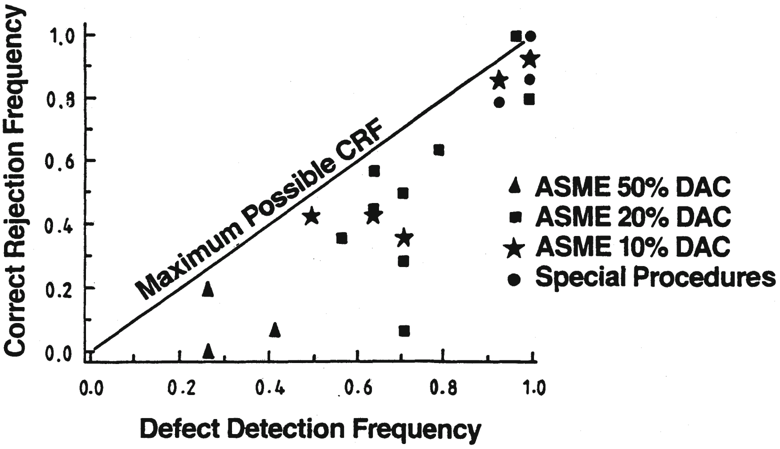

Fig. 3. Detection and Sizing Performance on Rejectable Flaws

In Figure 4, CRF is plotted as a function of RAF. This indicates that the three teams achieving 100% CRF did so with "inefficiencies" of between 40% and 70%, that is, that oversizing led to the rejection of 40% to 70% of the intended acceptable flaws in Plate 3. The presence of satellite reflectors may influence this aspect of the analysis and this is another area warranting further study.

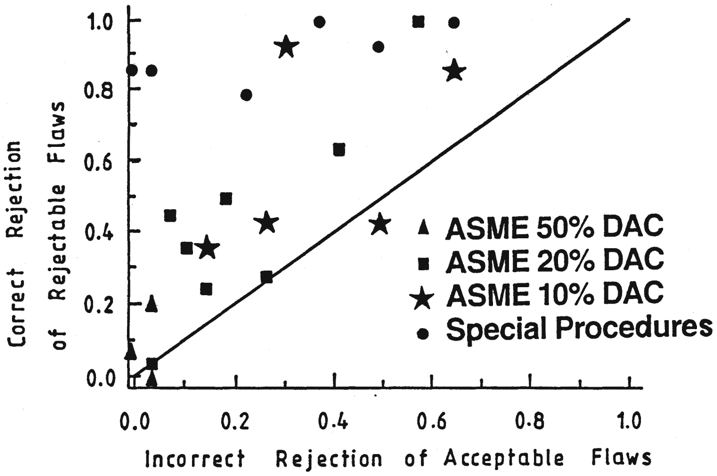

Fig. 4. Rejection of Acceptable Versus Rejectable Flaws

THE FORWARD PROGRAMME

It is planned to amplify the demonstrated effectiveness of SHERPA on practical inspection procedures for an operating nuclear power station in the CEGB.

The UKAEA at Risley Laboratories has developed a computer-based system for simulating, in real-time, ultrasonic inspection(5). This reproduces the important features and conditions of a manual inspection using stored ultrasonics data and obviates the need for a wide range of specific test blocks. The system will enable the calibration and scanning procedures under controlled conditions to be recorded, replayed and analysed, and a quantitative evaluation to be made of coupling efficiency, scan-coverage, measurement accuracy and data interpretation. The studies will be conducted with appropriate control of environmental conditions, and relevant operator characteristics will be recorded. In addition, the importance of operator characteristics is being investigated in collaboration with the Scottish School of NDT. This initially has taken the form of collecting relevant personnel details and views on certification tests, which are then related to operator performance. Over 100 subjects have been analysed to date. Similar information will become available from the PISC programme.

The retrospective analysis of the PISC II results will continue by discussion of the sources of error in detecting and sizing flaws in order to understand the detailed causes. In PISC III a substantial programme of reliability studies is planned to test and confirm the conclusions of the retrospective analysis. The influence of important environmental and psychological factors will be investigated by planned interaction within the framework of the various Actions. The programme for Action 2 at MPA Stuttgart is well advanced and is planned to commence in the last quarter of 1988.

CONCLUSIONS

Variability in defect detection is strongly influenced by operator error even when working to nominally similar inspection procedures. The identification of the causes of these errors enables appropriate safeguards to be instituted. Improved detection reliability can be achieved by more rigorous specification and application of both inspection and training procedures, using the SHERPA methodology.

The good results in the PISC II exercise were obtained under laboratory conditions, in many cases with highly qualified and well-motivated staff. The implications of transferring the inspection into field conditions with different inspectors and environmental conditions need to be investigated in order to provide confidence in inspection reliability in routine inspections. The PISC programme and the use of the computer simulator will contribute substantially to these studies.

The retrospective analysis has shown that for some teams the interpretation of ultrasonic sizing data is subject to substantial error. In these cases, significant flaws were detected but grossly undersized leading to the acceptance of rejectable flaws. In other instances, good rejection rates were achieved due to a significant degree of oversizing, but this also results in the rejection of many acceptable flaws. This has obvious consequences, including the expense of unnecessary repairs. The resolution and elimination of the causes of under- and over-sizing by these teams will contribute to a major improvement in their reliability and cost effectiveness.

ACKNOWLEDGEMENTS

The programme presented has been partially funded by CEGB under the Thermal Reactor Agreement. The contribution of H Seed (Risley Laboratory), S Crutzen and Mrs C Vinche (JPC Ispra) in performing the PISC II retrospective analysis is gratefully acknowledged. The contributions of K Stinson, B Leigh, R Berisford and G Aspin of CEGB in performing the evaluation of the effectiveness of SHERPA was most valuable, and the contribution of Human Reliability Associates during the development and application of SHERPA is acknowledged.

REFERENCES

1. Murgatroyd R A - "Human Factors in NDT". Seminar on Human Factors. European Safety, Reliability and Research Development Association, March 14 1988, Bournemouth, UK. (To be published)
2. Behravesh M M, Karimi S S, Ford M E - "Human Factors Affecting NDT Technician Performance". 9th International Conference on NDE in the Nuclear Industry, April 1988, Tokyo.
3. Murgatroyd R A, Crutzen S, Vinche C, Seed H - "Retrospective Analysis of PISC II Plate 3 Inspection: Results to March 1988". 9th International Conference on NDE in the Nuclear Industry, April 1988, Tokyo.
4. Murgatroyd R A, Embrey D, Ballard G, Tait J - "The Reliability of Ultrasonic Inspection". Proceedings of 8th International Conference on NDE in the Nuclear Industry, P83, 17-20 November 1986, Kissimee, USA.
5. Simm K - "Computer Simulated Flaws have put ultrasonic testers to the Test". Nuclear Engineering International, April 1988, pp43-44.