
An Evaluation of Human Factors Research for Ultrasonic Inservice Inspection

Manuscript Completed: February 1998
Date Published: March 1998

Prepared by
D.J. Pond, D.T. Donohoo, R. V. Harris, Jr.

DISCLAIMER

Pacific Northwest National Laboratory
Richland, WA 99352

D. Jackson, NRC Project Manager

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Prepared for
Division of Engineering Technology
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001
NRC Job Code W6275

dy
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER



DISCLAIMER

**Portions of this document may be illegible
electronic image products. Images are
produced from the best available original
document.**

Abstract

This work was undertaken to determine if human factors research has yielded information applicable to upgrading requirements in ASME Boiler and Pressure Vessel Code Section XI, improving methods and techniques in Section V, and/or suggesting relevant research. A preference was established for information and recommendations which have become accepted and standard practice.

Manual Ultrasonic Testing/Inservice Inspection (UT/ISI) is a complex task subject to influence by dozens of variables. This review frequently revealed equivocal findings regarding effects of environmental variables as well as repeated indications that inspection performance may be more, and more reliably, influenced by the workers' social environment, including managerial practices, than by other situational variables. Also of significance are each inspector's relevant knowledge, skills, and abilities, and determination of these is seen as a necessary first step in upgrading requirements, methods, and techniques as well as in focusing research in support of such programs. While understanding the effects and mediating mechanisms of the variables impacting inspection performance is a worthwhile pursuit for researchers, initial improvements in industrial UT/ISI performance may be achieved by implementing practices already known to mitigate the effects of potentially adverse conditions.

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

Contents

	Page
Abstract	iii
Executive Summary	vii
1 Introduction	1
2 PISC Overview	3
3 Variables Affecting UT/ISI	4
3.1 Complex Conditions	4
3.2 Heat	4
3.3 Noise	5
3.4 Task Duration	6
4 Fatigue, Self-Reports, and Performance	7
5 Individual Differences	8
6 Human Factors of UT/ISI Equipment	9
6.1 Organizational and Social Influences	10
7 Recommendations	11
7.1 Recommendations Regarding Future Research	11
7.2 Recommendations Regarding Current UT/ISI Practice	12
8 References	13
Appendix A - A Critical Review of the PISC III Action 7 Report No. 31, Human Reliability in Inspection	A.1
Appendix B - Updated Definitions for the Ability Categories in Recent Forms of the Manual for the Ability Requirements Scale (MARS)	B.1

Tables

2.1 Summary of the PISC III study variables	3
3.1 U.S. Department of Defense "limited tolerance" time limits at different temperatures	5

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

Executive Summary

Manual Ultrasonic Testing/Inservice Inspection (UT/ISI), which represents a significant proportion of the inspection in nuclear power plants, is a complex task requiring a combination of human sensory, perceptual, cognitive, and motor components. Unfortunately, inspection performance levels have often been found less than satisfactory, and substantial research has been undertaken to define the variables contributing to inspection performance problems and, then, to mitigate their influence. The variables given primary attention in the current review are derived from a research study which focused on UT/ISI--the "PISC III" study conducted by the Programme for the Inspection of Steel Components (PISC). These variables include task complexity, environmental factors, organizational and social issues, and a number of individual worker characteristics. Some of the accepted human factors guidance and recommendations relevant to UT/ISI performance are also summarized herein.

The scope and objectives of the present work are to:

- (1) determine if PISC III or other human factors research have yielded information which might be applicable to upgrading requirements in American Society of Mechanical Engineers (ASME) Boiler and Pressure Vessel Code Section XI and improving methods and techniques in ASME Section V; and
- (2) suggest research which might contribute to development of such upgrades and improvements.

UT/ISI represents a highly complex work situation. This complexity makes it a challenge to develop standards such as the ASME Codes as well as to design the research required to support Code requirements. For example, the extent to which the full panoply of environmental conditions is represented often distinguishes laboratory from industrial or other "real-world" research, and there are likely to be dramatic differences between the number and the influence of the environmental variables extant. In either circumstance, however, the combined effects of any pair of these variables can be *additive* (i.e., the total effect equals the sum of their individual effects), *synergistic* (i.e., the total effect is greater than the sum of their individual effects), or *antagonistic* (i.e., the total is less than the sum of their individual effects). Further, a combination that, for example, displays synergism on one measure might be antagonistic on another.

The complexity of the work situation and our limited ability to predict how variables will interact in real world situations contribute to difficulties in applying research results in this environment. Bailey (1996) has reported that "it seems the actual level of heat necessary to degrade performance varies from situation to situation and from individual to individual." Similarly, "research into the effects of noise on either mental or psychomotor performance has given very contradictory results: noise is just as likely to improve performance as to make it worse" (Grandjean 1985). Another variable yielding equivocal effects is task duration, for which conflicting evidence is reported from laboratory versus industrial research. Dissociation has also been found between self-reports of fatigue and measures of associated performance.

Gaining a comprehensive understanding of the effects and mediating mechanisms of the myriad variables impacting inspection performance is, of course, a worthwhile pursuit for researchers. On the other hand, making initial improvements to industrial UT/ISI performance may not require such a "basic" understanding but, rather, the simple implementation of practices already known to minimize the effects of potentially adverse conditions. For example, the Department of Defense has defined "limited tolerance" time limits for work in elevated temperatures, while NUREG 0700 Rev. 1 (1996) provides an indication of "stay times" for workers in elevated thermal conditions and also recommends training for workers in the

recognition and treatment of heat illnesses, water and salt replacement, and acclimatization mechanisms. Similarly, relevant human factors guidelines are available regarding high noise conditions (e.g., NUREG 0700) and equipment design (e.g., MIL-STD-1472).

While implementation of such guidelines and practices is, of course, advisable, it is important to note that Craig (1984) has stated "the problem is not so much of physical characteristics of the job...as of the social environment in which the work takes place." This social environment refers to interactions with co-workers as well as to managers and organizational policies. Spanner, Badalamente, Rankin, & Triggs (1986) note that performance is affected not only by managerial policies, but also attitudes and expectations which have been found to "influence an inspector and have a marked effect on the relative frequency of errors, even if the inspector is not responding to direct management instruction." Finally, Karimi (1988) reports that the "...contextual factors (e.g., supervisory and ISI managerial practices) are the major determinants of the outcome of performance...[and] industry may need to place a greater emphasis on altering the existing philosophies, attitudes, and practices of nuclear power plant supervisors and managers."

Also of potential significance to UT/ISI performance are "individual differences" such as personality and, especially, each inspector's relevant Knowledge, Skills, and Abilities (KSAs). For example, Harris & McCloskey (1990) found that the presence of four cognitive factors--development and testing of explicit hypotheses; avoidance of premature conclusions; application of if-then logic; and not disregarding evidence--was associated with a 400% improvement in inspection success rate relative to those inspections in which these factors were not in evidence. Similar research in the other aspects of UT/ISI performance may be justified as well. A useful reference in this regard is Fleishman's *Manual for the Ability Requirements Scales* (see Appendix B) which defines 52 human abilities, a number of which--e.g., perceptual speed, deductive reasoning, control precision, problem sensitivity--appear likely to be required for effective UT/ISI performance.

Recommendations Regarding Future Research

Overall, the priority should be on research which has a primary objective to develop strategies for mitigating UT/ISI performance decrements rather than to understand the mediating factors which contribute to these decrements. That is, a shift in research emphasis away from such variables as effects of heat and noise--which industry has little ability to control--and towards those such as managerial policies, job design, and defining UT/ISI-relevant KSAs. The approach should be to build on what is known in industry about avoiding performance decrements rather than artificially creating such decrements in order to resolve equivocal research findings. To assure the greatest return on research investments, research teams should comprise those knowledgeable in UT/ISI as well as those with expertise in research design, conduct, and analysis. Finally, as above, task analysis results will provide an important knowledge base on which to establish meaningful research programs.

Recommendations Regarding Current UT/ISI Practice

The present work enables a number of recommendations to be made regarding the current practice of UT/ISI in industry. First is to adopt or adapt available relevant standards/guidelines for such issues as "stay times" (i.e., work-rest cycles) and training for workers in elevated thermal environments; use of ear defenders in high-noise conditions; and incorporation of human factors principles in equipment and job design. Second is to require certification, performance demonstration, training, and research to be conducted in conditions which resemble the field as much as possible in their physical and social aspects. Third is to utilize the skills of organizational effectiveness specialists in order to assess and, as necessary, help to modify the philosophies, attitudes, and practices of nuclear power plant supervisors and managers, and to encourage the adoption of managerial practices which contribute to mitigating potential performance decrements. Last, as a foundation for many of the above activities, require task analyses of UT/ISI operations and determine the KSAs required for successful inspection performance.

1 Introduction

Inspection task performance, long an object of human factors research, is known to be subject to influence by dozens of variables (see Wiener 1984 for a review of studies dating back to the 1940's), and more than five dozen such variables have been catalogued by Megaw (1978). Ultrasonic testing/in-service inspection (UT/ISI) is one of a number of approaches to non-destructive inspection (NDI)¹ utilized to search for flaws such as intergranular stress-corrosion cracking (IGSCC) in welds or other materials. UT/ISI is a complex task comprising a combination of sensory, perceptual, cognitive, and motor components. Within each of these, application of multiple skills is required for effective performance and Harris and McCloskey (1990), for example, have discerned nine cognitive (or information processing) elements of ultrasonic inspection which are individually and collectively correlated with inspection success.

Focusing specifically on UT/ISI, Triggs, Rankin, Badalamente, & Spanner (1986) defined five classes of factors known to affect performance: task, procedural, training, individual, and environmental variables and, in a companion volume, provided a review of research on variables in each of these categories (Spanner, Badalamente, Rankin, & Triggs 1986). While a detailed exposition of these findings is beyond the scope of the present discussion, it is clear that UT/ISI represents a highly complex work situation and that this complexity makes it a challenge to develop standards such as the American Society of Mechanical Engineers (ASME) Codes as well as to design the research required to support Code requirements.

Craig (1984) has noted "there is more than a suspicion that real task demands and influences have not been (or cannot be) replicated in an experimental or laboratory context." Indeed, many of the variables found to influence inspection performance have only done so in laboratory tests and/or in military studies (Eastman Kodak 1983). The relevance of such laboratory-based research findings, or those derived from studies employing military subjects, to performance in industry may be questioned due to differences in such variables as context, motivation, age,

and physical fitness, among others. Further, Moray (1991) notes that "much of the vast literature on human attention...is irrelevant to the study and understanding of large scale industrial systems," and Behraves, Karimi, & Ford (1989) report that nuclear power plant inspectors often exhibit good performance during training but poor performance in the field. In light of these circumstances, the Programme for the Inspection of Steel Components (PISC) and, in particular, their "PISC III" study and report (1994) attempted to address some of the key UT/ISI performance issues by "studying and identifying causes of variability in inspection activities, and identifying some of the factors influencing the reliability of inspection in industrial conditions." While an overview of this study is presented below, a "Critical Review" has already been prepared and delivered (Doctor 1995, see Appendix A), and replication of this previous report will be specifically avoided.

The scope and objectives of the present work are to

- (1) determine if PISC III or other human factors research have yielded information which might be applicable to upgrading requirements in ASME Section XI and improving methods and techniques in ASME Section V
- (2) suggest research which might contribute to development of such upgrades and improvements.

In reviewing the literature related to the above objectives, where possible a preference was established for information and recommendations which have become accepted and standard practice as opposed to those which are "cutting edge," and often unreplicated, research findings. In turn, this typically led to acquisition of information in such secondary sources as textbooks, handbooks, guideline documents, review articles, and other compilations. This approach was adopted with the expectation that such information is more likely to enhance the relevance and immediate applicability of recommendations made herein to current practice as well as to future studies. This report is not intended as a comprehensive review of the literature nor was there any attempt to include all UT/ISI publications of a particular researcher (e.g., D.H. Harris) or organization (e.g., Pacific Northwest National Laboratory).

¹Also known as non-destructive testing (NDT) and non-destructive evaluation (NDE).

Instead, references selected for inclusion are by-and-large those which were considered to best represent the findings/conclusions extracted from the set of documents reviewed relative to each aspect of UT/ISI.

The Code of Federal Regulations requires that reactor components be inspected on a periodic basis. While a number of automated UT/ISI systems have been developed, situational constraints in typical nuclear power plant environments--for example, adequate clearance for sensors/probes--frequently make use of such devices infeasible. As a result, manual UT/ISI currently represents a significant proportion of the inspection in nuclear power plants and it is likely to do so for the foreseeable future. A number of studies have shown that the inspector is the key element determining overall UT/ISI effectiveness (see, e.g., the PISC III Report 1994, and Wheeler, Rankin, Spanner, Badalamente, & Taylor 1986) as wide variations in performance have been reported for personnel utilizing the same equipment and following the same procedures. For these reasons, the focus of this report is on manual UT/ISI².

As summarized by Harris (1988), manual UT/ISI comprises the following functions

- determination of the inspection strategy
- acquisition of the historical and geometric information about the weld area
- preparation and calibration of the inspection equipment
- manual control of the scanning pattern and movement of the transducer
- detection of defect indications by recognizing and interpreting signal patterns
- selection and application of appropriate defect detection procedures
- discrimination of defect signals from various possible sources

²It is, however, worth noting that the human factors issues associated with automated UT/ISI are likely to differ somewhat from those with manual UT/ISI, and are worthy of review in their own right.

- creation of records of inspection data, methods, and results.

These flaw detection procedures often take place in a radioactive environment and require use of appropriate protective garb. The inspector holds a relatively small transducer (see Spanner, et al. 1986 for typical transducers) which is connected by wire to an oscilloscope-type CRT display unit. This transducer is moved in prescribed scanning paths over the surface of the material being inspected--e.g., a welded pipe joint--and the oscilloscopic screen is simultaneously monitored for displayed patterns which are indicative of flaws.

As noted above, UT/ISI performance requires a combination of sensory, perceptual, cognitive, and motor skills and, of course, is subject to human (inspector) error or *system-induced human error* (i.e., having antecedent situational contributors) in any of these processes³. Inspection performance levels have often been found less than satisfactory, and substantial research, including much human factors research, has been undertaken to define the variables contributing to inspection performance problems. Such definition is a necessary precursor to understanding the mechanisms through which these variables work and, finally, to devising strategies which mitigate their impact.

Industrial inspection tasks are often characterized as *vigilance* tasks (see, e.g., Parasuraman 1986), and the same theory and evaluation methods--*signal detection theory* and the *receiver operating characteristic*--are applicable to each (see, e.g., Boff & Lincoln 1988). Further, the nature of inspection failures is seen as analogous to time-related *vigilance decrements* in performance (see, e.g., Warm 1984). However, many differences between typical vigilance and inspection tasks have been noted (see Wiener 1984 for a detailed discussion) and the extent to which research findings in the former circumstance apply to the latter may be questioned. Despite these differences, much of the current knowledge about variables which influence inspection performance is derived from traditional vigilance-type tasks.

³Although a detailed discussion of errors is beyond the scope of this review, many excellent sources are available. See, for example, Park (1997), Dhillon (1986) and, especially, the collection of papers by James Reason and others in Rasmussen, Duncan, & Leplat (1987) for general discussions of human error. Additionally, Webster (1989) provides specific consideration of errors in UT.

Therefore, UT/ISI researchers must base the selection and level of their independent variables--that is, those (e.g., heat) which are manipulated to determine the effect on some dependent variable (e.g., performance)--on such research. The present discussion is largely based on such research, and therefore subject to the same shortcomings, as well.

2 PISC III Overview

One attempt to study UT/ISI performance under conditions approximating those in operational environments was the PISC III study on human reliability. These researchers developed "typical inspection tasks" through which to assess the simultaneous influence of a number of variables suspected of impacting inspection performance in operational conditions. As summarized in Table 2.1, the PISC III study required the inspectors to perform a variety of UT/ISI tasks (e.g., calibration, scanning, maximizing flaw indications, plotting) under "laboratory" and "simulated industrial" conditions. Additionally, in the PISC III report reference was made to a lower level of illumination in the simulated industrial inspection environment, but no further specification was provided.

The PISC III researchers utilized a battery of UT/ISI performance measures (e.g., flaw detection frequency, plotting accuracy) as well as a variety of "psychological" tests (e.g., an introversion-extraversion inventory, self-reports of tiredness) and physiological measures (deep body temperature, pulse rate).

Additional details regarding the PISC III study may be found in Appendix A and, for a comprehensive presentation, in the PISC III Report itself (1994). For present purposes it will suffice to note that although the earlier review and evaluation (Appendix A) found much to be praised about the PISC III methodology, there were a number of shortcomings noted. Some of these were seen as critical (e.g., abandoning select measures in mid-experiment), thereby diminishing the study's potential utility and impact. Indeed, in some instances the conclusions drawn (e.g., the impact of "tiredness" on inspection performance) were found to be unsubstantiated by the details presented. In PISC III the environmental and other variables were presented in combined fashion (i.e., "as a set") and the research design limited the ability to assess the contribution of each variable, as well as combinations of variables, to inspection performance or other reported effects. To gain a clearer understanding of the influence of the industrial UT/ISI environment on inspection performance, it is necessary to use research designs and statistical analyses which enable such effects to be partitioned and assessed.

Table 2.1 Summary of PISC III study variables (adapted from the PISC III Report 1994)

Variable	Laboratory conditions	Simulated industrial conditions
Shift length	8 hrs	12 hrs
Shift times	0830-1630	0830-2030
Breaks	1 x 1 hr 2 x 15 min	1 x 1 hr 1 x 45 min; 2 x 15 min
No. of work days	5	11
No. of rest days	0	1 (after work day 6)
Temperature	64-73°F	77-81°F
Humidity	Unspecified ambient	40-50%
Noise level	Unspecified "quiet"	80 dB(A) recording of industrial noise

3 Variables Affecting UT/ISI

3.1 Complex Conditions

The extent to which the full panoply of environmental conditions is represented often distinguishes laboratory from industrial or other "real-world" research. Of course, even when earnest attempts are made to simulate industrial situations, there are likely to be dramatic differences between the number and the influence of the environmental variables extant under these controlled circumstances and those present in true operational situations. For example, thermal conditions, one of the most-studied environmental influences on human performance, comprise a complex set of variables. Different kinds of work are differently affected by heat (see, e.g., Boff & Lincoln 1988), the effects of heat may differ considerably in noisy and quiet environments (Parasuraman 1986) and, importantly, the effects of heat are not always detrimental (see, e.g., Hockey 1986).

Of course, these kinds of relationships are further complicated when other variables such as presence of co-workers or duration of task are added. Each of these may have one kind of effect when studied singly and another, often more complex and unpredictable effect when presented in combination. Such effects can be quite complex, with any pair of variables combining *additively* (i.e., the total effect equals the sum of their individual effects), *synergistically* (i.e., the total effect is greater than the sum of their individual effects), or *antagonistically* (i.e., the total is less than the sum of their individual effects). Further, Murray & McCally (1973) note that "a combination of stresses that is synergistic by one criterion might be antagonistic by another." In addition to an excellent discussion and detailing of these general phenomena, Murray & McCally (1973) also note:

A number of obvious factors have limited the study of combined environment stress in spite of the obvious operational importance of such work. First, experimental design is exceedingly complex...Second, this type of research requires complex experimental facilities not generally available. Finally, to assure valid conclusions, large number of subjects are required.

3.2 Heat

Bailey (1996) among others has reported that "it seems the actual level of heat necessary to degrade performance varies from situation to situation and from individual to individual." Although PISC III did assess different aspects of inspection performance (calibration, scanning, plotting, and so forth), the levels of heat chosen may have not been appropriate to impact performance on these kinds of tasks. As noted in Table 2.1, the PISC III study compared UT/ISI performance under two thermal conditions which ranged between 64-73°F and 77-81°F, respectively, with only minimal effects resulting. This is not particularly surprising since, for example, Echeverria, Barnes, and Bittner (1991) found vigilance performance decrements to begin at 86-92°F based on their review of over 600 research studies involving heat and other environmental variables. Similarly, Hancock (1984) has noted that "the current weight of empirical evidence suggests that vigilance performance is impaired by exposure to temperatures greater than 32°C (89.6°F) ET [Effective Temperature⁴]." Indeed, some research has found 26°C (78.8°F; roughly the mid-point of the PISC III high-heat range) to be optimal for vigilance performance (see Hancock 1984 for a discussion). The heat levels chosen for PISC III are also puzzling in light of the fact that "many UT/ISI tasks take place in elevated temperature environments where 90°F and high humidity conditions are common." (Spanner, et al. 1986).

The PISC III researchers note considerable differences among inspectors' self-reports of impacts associated with the different thermal conditions, and they also report differences between the American and European participants in this regard. Such individual differences/preferences--of heat as well as other variables--are potentially important sources of variation in research as well as in real-world UT/ISI performance. The PISC III protocol did not, however, enable measurement of differences in, for example, self-reported "tiredness" associated with these different thermal conditions, and it will be important in future studies to employ research design and analyses

⁴Effective Temperature (ET) is a measure which incorporates air temperature, relative humidity, and air velocity. In many cases, ET is likely to be higher than the air temperature in conditions of high humidity and lower than air temperature when the air is moving at high velocity.

which enable the impact of such differences to be accounted for in interpreting performance, psychological, and physiological data.

Spanner et al. (1986) note that inspectors typically must wear protective clothing and masks, each of which may contribute to heat stress. Of course, the "effects of heat may become very severe in situations where workers are required to wear heavy protective clothing or where employees are exposed to several exposures at one time such as...in noisy and poorly illuminated areas" (Echeverria, et al. 1991). Inclusion of protective clothing and masks in the PISC III research would not only have more closely replicated typical industrial conditions, it may also have increased the inspectors' thermal load to levels at which more obvious performance effects may have resulted.

While, as noted, there are specialized circumstances in which exposure to high heat has been associated with enhanced performance--(e.g., experienced subjects on initial exposure--see, e.g., Hancock 1984 for a discussion), these do not appear to be generally applicable to industrial UT/ISI performance. And while further research into thermal effects should be continued, as a general practice it may be advisable to simply strive to maintain the thermal conditions at some moderate level. The U.S. Department of Defense (DoD; MIL-STD-1472) has defined "limited tolerance" time limits for work at the temperatures shown in Table 3.1.

Table 3.1 U.S. Department of Defense "limited tolerance" time limits at different temperatures (adapted from MIL-STD-1472)

Temperature (°F)	Time of limited tolerance (hours)
102.9	1/2
99.9	1
95.0	2
89.9	4
84.9	12

These ambient temperature values do not reflect use of protective clothing nor do they account for different levels of activity associated with different tasks. NUREG 0700 Rev. 1 (1996) indicates that "the level of physical activity and required protective clothing, as well as temperature and humidity, should be considered when assessing the danger of heat exposure posed to operators." The resulting "stay times" for workers in, for example, 93°F heat varies from as little as 10 minutes to as much as 8 hours depending on the protective clothing worn (i.e., on the restriction of evaporative heat loss) and worker metabolism (i.e., the amount of metabolic heat produced as a consequence of the work being performed). This range narrows to between 20 minutes and 4 hours as one considers an inspector wearing single or double layer coveralls and performing UT/ISI tasks which represent a low to moderate metabolic load. In addition to the work-rest cycles thus defined, NUREG 0700 Rev. 1 (1996) also recommends training workers in the recognition and treatment of heat illnesses, water and salt replacement, and acclimatization mechanisms.

3.3 Noise

"Research into the effects of noise on either mental or psychomotor performance has given very contradictory results: noise is just as likely to improve performance as to make it worse" (Grandjean 1985). Similar to the preceding discussion, the available research would generally not predict performance impacts at the level of noise employed in PISC III, 80 dB(A).⁵

Hancock (1984) has reviewed a number of studies in which continuous (typically, white or pink)⁶ noise did have an impact on vigilance performance, but these effects appeared at levels of 90 to 100 dB, with one investigation reporting "slightly depressed performance" at 85 dB. Hancock (1984) summarized a 1979 study by Thackray and Touchstone which is somewhat analogous to UT/ISI and the PISC III research and in which the effects of 57 dB versus 78-80 dB noise recorded in operational air traffic control rooms were investigated. As reported by

⁵dB(A) refers to an "A weighting" system which adjusts the measure to reflect the relative sensitivity of humans across the auditory spectrum; that is, a generally lessened sensitivity to lower frequencies.

⁶ White noise is that in which all audible frequencies are equally represented. Pink noise is an approximation of white noise. Both types resemble the static heard when a radio is tuned between stations.

Hancock, neither of these levels had an impact on detection and interpretation of signals on a radar screen for either low- or high- complexity tasks.

Summarizing the work of others, Proctor & Van Zandt (1994) detail five kinds of tasks which have been associated with performance decrements during exposure to noise levels as low as 80 dB(C).⁷ These are tasks which:

- are of extended duration
- require a steady gaze and might be disrupted by startling noises
- are unimportant
- require comprehension of verbal material
- are largely unspecified and may require a rapid change of response. Of these, only the first might be construed to characterize the typical industrial UT/ISI task situation, although, of course, an individual inspector might consider the task unimportant, particularly so in non-critical research situations.

Still, 80 dB(A) noise may constitute an appropriate level to be used in research if it is representative of the industrial UT/ISI environment and its effects are to be assessed under the combined influence of other variables. Of course, as noted, the dependent measure(s) employed is also critical (see, e.g., Boff & Lincoln 1988). In PISC III, for example, although noise effects were not evident in the performance measures, different behavioral responses were reported as "most of the inspectors were conscious of the noise background, some more than others, and many chose to wear the ear muffs⁸ that were available." Future research needs to be designed to partition and analyze such results.

In parallel with such research we must continue to address the potential impact of noise on UT/ISI performance by mitigating exposures in accordance with, for example, NUREG 0700 Rev. 1 (1996) which states "ear protection

devices should be available and required to be worn in areas where noise levels are 85 dB or more." This, presumably, is for hearing protection as are levels specified by the Occupational Safety and Health Administration (which, virtually by definition, is concerned with protection from hearing damage rather than with maintenance of performance). With respect to "non-hazardous noise", the DoD stipulates in MIL-STD-1472 that general workspaces such as maintenance shops and other areas requiring occasional direct communication at distances up to five feet--akin, for example, to communications between UT/ISI inspectors as they coordinate their activities--shall have noise levels not greater than 75 dB(A).

3.4 Task Duration

It seems logical to assume that beyond a certain duration performance begins to decline and, to some degree, this is supported by research. For example, Proctor & Van Zandt (1994) note that "...working additional hours usually results in a decreased rate of productivity relative to that shown for the standard 40-hr. week." Similarly, Tepas & Monk (1987) conclude that "common sense and the literature suggest that lengthening the workday for some jobs may increase worker errors and worker accidents. Jobs that are very boring and/or require heavy physical work are probably not good candidates for the extended workday." In a laboratory study, Mahan (1990) found that a task requiring "judgement" and integration of information--each of which are also required for UT/ISI (see, e.g., Harris 1988)--was more prone to degradation over a four hour period than were logical reasoning and mental arithmetic tasks (each termed "analytical").

Of course, the extent to which UT/ISI and other vigilance/inspection tasks represent these conditions may be questioned. For example, after reviewing the literature on inspection performance, Wiener (1984) notes that although many researchers refer to time-induced decrements in performance, and others "uncritically" accept the existence of time decrements, "the literature simply does not indicate whether or not time decrements occur on the shop floor."

Among the possible reasons he offers for this is that while vigilance researchers have learned how to create task situations which induce time decrements, industrial workers and managers have learned how to avoid these situations. Parasuraman (1986) surmises, for example, that

⁷dB(C) provides no adjustment for humans' relative sensitivity.

⁸Ear "muffs" refer to what in America are typically referred to as ear plugs or other forms of ear defenders. That is, those which afford protection against auditory rather than thermal conditions.

such management practices as work-rest scheduling, task rotation, and worker self-pacing are likely to reduce potential performance decrements. In terms of research applicability, one might question whether the PISC III inspectors adhered to a break schedule similar to that imposed during the research (see Table 1.1) when working at their actual jobs, and whether they take might also take "informal" breaks from their work when not being continually observed by the research monitors.

In an applied study of simulated control room operations, Baker (1996) found that "neither mental performance nor objectively measured alertness were significantly impaired on 12-hr night shifts as compared with 8-hr night shifts. Subjective alertness and mood were, for the most part, not negatively affected by the extended work shifts." On the other hand, the PISC III study compared performance of inspectors working 8- versus 12- hour shifts beginning at 0830 hours and concluded that "working long shifts...can cause tiredness and demotivation for some inspectors." While this conclusion may be correct, because these sessions were all conducted on successive days the inspectors on the 12-hour task not only performed an additional 3 hours 15 minutes (i.e., 4 hours minus one 45 minute break) per day, they also had 4 hours less rest between work sessions. They were also performing later into the evening and, as a consequence, likely at different points in their circadian rhythms for body temperature, performance efficiency, and other factors (see, e.g., Parasuraman 1983). Despite the fact that these work schedules may accurately reflect those encountered in industry, for research purposes each of the additional differences between the 8- and 12-hour tasks served to confound the experimental situation, thereby lessening the certainty with which valid conclusions could be drawn from the results. Latack & Foster (1985) found an improvement in information systems operation performance (i.e., a decrease in operator-caused errors) over the course of the first year following implementation of a 3 day-38 hours/week work regimen. However, a "preliminary investigation" of overtime operations by Baker, Olson, & Montgomery (1990) revealed that a 12-hour operator shift schedule was significantly related to a higher incidence of operator-error Licensee Event Reports. Conceivably, then, it may not be the hours per day which is the critical variable in influencing performance but, rather, the hours per week or month and/or the amount of rest over this period which

impact performance. Consequently, careful and appropriate scheduling may contribute to maintaining effective UT/ISI performance, in part because "as the number of consecutive days worked decreases and/or as the number of consecutive days available to rest increases, the onset of stress and fatigue is arrested" (Pierce & Dunham 1992).

Tepas (1985) provides an extensive list of advantages and disadvantages of compressed workweek schedules and, additionally, reports there are few experimental studies of workers with actual "compressed workweek" experience-- i.e., those which involve workdays of, for example, 10 or 12 hours per day in a 40 hour workweek. He also summarizes the work of others which concluded that there is little to suggest that fatigue is a serious problem in these situations. Nonetheless, it is clearly difficult to draw firm conclusions regarding the impact of task duration on fatigue and, by extension, on UT/ISI performance. While well-conceived and properly conducted research should, of course, continue, in operational situations it is probably advisable to adopt work schedules and practices (e.g., job rotation, inspector-determined breaks, etc.) which are likely to mitigate potential adverse effects of extended duration performance.

4 Fatigue, Self-Reports, and Performance

PISC III researchers assessed "tiredness" through inspector self-reports. While this is a common and viable strategy, it is not without shortcomings. For example, Pierce, Newstrom, Dunham, and Barber (1989) note that compressed workweek schedules are consistently associated with worker reports of fatigue. However, they further note the possibility that workers *expect* to be more fatigued under such schedules and, when asked, simply confirm their own expectations. Also problematic are the different variables which mediate this task duration-performance relationship and the lack of commonly-accepted definitions for these variables (see, e.g., Hockey 1986). There may be, for example, difficulty in effectively distinguishing between such states as *tiredness* and *fatigue* as well as between *muscular* versus *mental fatigue*. In traditional vigilance tasks as well as in many industrial jobs, the

related concept of *boredom* also enters the equation. More comprehensively, Luczak & Mueller (1994) note: Fatigue and similar states of feeling tired, bored, and so on in working life appear in different forms, such as

- fatigue as an immediate consequence of work
- fatigue as a result of deficiency of sleep
- fatigue-like states related to diurnal variations
- fatigue-like states resulting from boredom
- decrement of attention on detection tasks (problems of vigilance).

Wheeler et al. (1986) found no difference in inspection performance between morning and afternoon sessions despite the fact that inspectors felt that fatigue did affect their performance. As previously indicated, Parasuraman (1986) suggests that self-paced inspection tasks may be more resilient to time-related performance decrements. Although he also notes that such performance effects are somewhat equivocal, he nonetheless recommends self-pacing "as it tends to reduce worker complaints of fatigue and boredom." There is, therefore, clearly the potential for significant incongruity between self-reports of tiredness and measures of inspection performance, and care must be taken in designing and assessing research in this area.

5 Individual Differences

PISC III incorporated one of the most widely-studied personality dimensions influencing vigilance performance, introversion-extraversion. While this is an often-used, useful, and apparently significant variable in vigilance research, there are a number of others that should be considered for inclusion in future studies. These additional vigilance-related personality dimensions include (see e.g., Berch & Kanter 1984)

- field dependence/independence: the relative tendency to be influenced by cues in the perceptual field

- locus of control: the extent to which one believes that rewards are contingent primarily on their own actions versus on external forces
- personality type: preference for a rushed, competitive lifestyle versus one which is more relaxed
- achievement motivation: orientation on success versus avoidance-of-failure.

In addition to personality variables, operator skills and abilities (including both inherent and learned aspects) comprise a set of variables critical to determining an individual's level of performance. For example, it was previously noted that UT/ISI performance comprises sensory, perceptual, cognitive, and motor aspects. These, in turn, are influenced by the individual's levels of ability and skill in each of these aspects as well as by the extent of his/her knowledge about the task. An understanding of the Knowledge, Skills, and Abilities (KSAs) associated with the various tasks comprising UT/ISI is, therefore, essential to the study of inspection performance, and studies not utilizing such information (e.g., PISC III) are handicapped in their ability to reach diagnostic and predictive conclusions.

Additionally, as detailed earlier, Harris & McCloskey (1990) have discovered nine cognitive factors contributing to UT/ISI performance. They found that the presence of four of these--development and testing of explicit hypotheses; avoidance of premature conclusions; application of if-then logic; and not disregarding evidence--was associated with a 400% improvement in inspection success rate relative to those inspections in which these factors were not in evidence. In conjunction with the Electric Power Research Institute (EPRI), Harris (1996) is continuing this research in order to guide inspector selection, certification, and training programs. Similar research in the other aspects of UT/ISI performance may be justified as well. A useful reference in this regard is *Taxonomies of Human Performance* (Fleishman & Quaintance 1984) which details numerous categorization schemes through which to assess human abilities and performance. Included among these is Fleishman's *Manual for the Ability Requirements Scales* (MARS; see Appendix B) which defines 52 human abilities, a number of which--e.g., perceptual speed, deductive reasoning, control precision, problem sensitivity--may be important for effective UT/ISI performance.

In addition to the relationships between various psychological traits and vigilance performance, an array of physiological indices has been investigated. The PISC III research initially included measures of inspectors' pulse rate and deep body temperature, and although an electrocardiogram was also utilized, the measures of cardiac function (other than pulse rate) to be considered were unspecified in the report. Because there was little response on these indicators by the first few inspectors, physiological measurement was abandoned. Nonetheless, other studies have demonstrated physiological responsiveness to vigilance task situations.

Measures which have been found to be indicators of weariness or fatigue include electroencephalograms (EEGs, a measure of the electrical activity of the brain), as indicated by increasingly synchronized brain waves, and critical flicker fusion (CFF) frequency, as indicated by a lowering of the frequency at which a flashing light appears to be continuous (Grandjean 1985). EEGs have been said to provide the most sensitive indicator of changes in vigilance. Autonomic system indices such as mean heart rate and skin conductance have been found to decline during a vigilance task, but these changes do not appear to be correlated with level of performance (Parasuraman 1983). Heart rate variability, on the other hand, increases during vigilance and is moderately correlated with performance (Parasuraman 1984). Given the specialized hardware requirements and the generally intrusive nature of most physiological measurement techniques (e.g., as noted in the PISC III report), most of the vigilance research involving such measures has been laboratory based, and their utility in applied situations such as UT/ISI appears to be quite limited. For the same reasons, physiological indices might be considered a lower priority for PISC-like research than other dependent measures.

6 Human Factors of UT/ISI Equipment

Heasler, Taylor, Spanner, Doctor, and Deffenbaugh (1990) have noted a general lack of design consideration for use of equipment in radiation-contaminated environments. Wheeler, et al. (1986) conducted a human factors evaluation of a number of ultrasonic testing devices and developed a checklist through which to evaluate the key

inspector-equipment interfaces. They note that most designs reflect a compromise between ideal operator interfaces and the need to provide an easily portable device.

As previously noted, one limitation of automated UT/ISI equipment is the additional clearance requirements typically associated with such devices. On the other hand, while manual devices--especially including search units--are often quite small (see, e.g., Spanner, et al. 1986), these may create human factors problems because they are difficult to hold and control, and often lack grip aids (Heasler, et al. 1990). In addition to the likelihood of contributing to scanning performance problems, these small units are most commonly held between the thumb and index fingers (Spanner, et al. 1986) in an undesirable "pinch grip" configuration which is known to contribute to repetitive stress injuries.

Despite the aforementioned compromises, however, none of the eight devices analyzed by Wheeler, et al. (1986) was deemed "unacceptable" from a human factors perspective. Although it will not be useful to fully detail the outcome of this assessment, a brief summary of key findings may be instructive. The primary control problems reported were:

- control size and/or spacing were inappropriate for operation by gloved inspectors
- control design and/or orientation didn't minimize or preclude accidental activation
- control scales were small and difficult to read.

The major display concerns included:

- ineffective control of glare on display screens
- no provision to adjust screen brightness, contrast, or focus
- small screen size, which made signal discrimination difficult.

Numerous human factors handbooks, guideline documents, and checklists are readily available to support the design and/or evaluation of equipment (see, e.g., Woodson,

Tillman, & Tillman 1992; MIL-STD-1472D 1989; Sanders & McCormick 1987; Grandjean 1985; Eastman Kodak 1986, 1983). The NRC also has relevant guidelines—for example, glare should not interfere with the readability of displays (NUREG 0700, Rev. 1 1996)—which could/should be applied to UT/ISI equipment. No attempt will be made here to provide a recitation of these guidelines. For present purposes, it will suffice to note that among the most important human factors principles related to UT/ISI equipment are to

- understand and design to accommodate the capabilities, limitations, and expectations of all anticipated users
- design the human-system interfaces to be easily, accurately, and effectively used and to minimize the consequences of misuse
- assure the adequacy of the design for use through the full life cycle (manufacture to disposal) of the equipment
- make these interfaces as “intuitive” as possible to minimize the need for specialized capabilities, training, job aids, and other support devices required to overcome inadequate care in design
- minimize the likelihood and consequences of system-induced human error.

Finally, although application of such human factors principles and guidelines might adequately address most inspection equipment interface issues, in recognition of the unusual aspects of UT/ISI work (operating in contaminated environments; use of protective clothing/masks; critical nature of inspection performance effectiveness, etc.), Wheeler, et al. (1986) have recommended development of specific design criteria based on a detailed task analysis conducted for field conditions. NUREG 0700 Rev. 1 (1996) provides guidance regarding the conduct of such task analytic efforts.

6.1 Organizational and Social Influences

The challenge of conducting laboratory studies which are relevant to industrial operation has been repeatedly

referred to herein. With specific reference to vigilance performance, Craig (1984) has stated “the problem is not so much of physical characteristics of the job...as of the social environment in which the work takes place.” For just this reason, Karimi (1988) has called for development of training conditions which “resemble the field as much as possible in terms of physical conditions as well as the social influences.”

This social environment refers to interactions with co-workers as well as to managers (see, e.g., Triggs, et al. 1986) and organizational policies. With specific reference to nuclear power plant inspections, Behraves et al. (1989) found that “contextual” factors—including managerial rules, standards, cooperation, and support, and the degree of organizational bureaucracy—were the most salient factors influencing inspection performance. Similarly, Spanner, et al. (1986) note that performance is affected not only by managerial policies, but also attitudes and expectations which have been found to “influence an inspector and have a marked effect on the relative frequency of errors, even if the inspector is not responding to direct management instruction.” They further report that in industrial practice inspectors have come under pressure to modify their decision criteria (i.e., decision to record an indication as a flaw), while co-workers may seek leniency during inspection of welds made by a fellow worker.

How can performance be improved? The...contextual factors (e.g., supervisory and ISI managerial practices) are the major determinants of the outcome of performance...[I]ndustry may need to place a greater emphasis on altering the existing philosophies, attitudes, and practices of nuclear power plant supervisors and managers (Karimi 1988)

This, then, would appear to be an ideal area for development of guidelines and, as required, the applied research necessary for their definition and support. The contributions of organizational effectiveness specialists, comprising, for example, those trained in industrial/organizational psychology and management science, may be particularly useful in these efforts.

The PISC III approach to “studying and identifying causes of variability in inspection activities, and identifying some of the factors influencing the reliability of inspection in industrial conditions,” afforded little consideration of social variables. In particular, the presence of evaluative

others--i.e., the research monitors who managed the test program and assessed the subjects' inspection performance--is likely to have had an influence on inspector performance and other responses. Further, the researchers attribute significant performance impacts-- "a major influence on the behavior and reliability"--to the presumed "isolation" of the simulated industrial inspection environment. It is not at all clear, however, the extent to which isolation is typically a part of actual inspection environments, nor if a sense of isolation is even achievable under evaluative circumstances such as those in PISC III, whether the evaluative others are physically and obviously present or not.

Motivation to perform is a critical variable in most tasks and it represents yet another significant difference between typical laboratory conditions and those evident in industry. It has direct impacts on performance and is also a cross-cutting variable which may interact with other variables--e.g., task duration, managerial practices, and individual characteristics like personality--in influencing UT/ISI performance.

Differences in locus of control (i.e., primarily influenced by internal vs. external sources) relate to such characteristics as being a self-starter vs. someone requiring repeated direction, and someone who does a good job primarily because of the satisfaction of so doing rather than for monetary or other external rewards. In operational conditions, the management approach (i.e., motivating influence) required to maintain a high level of inspection performance by these different types of individuals would clearly need to be different. This is true at the individual level as well as at the "macro" or cultural level. For example, Monden (1993) reports distinct stereotypical differences between the American focus on monetary rewards versus the Japanese focus on job satisfaction and feelings of accomplishment.

There are many ways to influence an individual's level of motivation. For example, the simple use of different characterizations of vigilance tasks--e.g., as monotonous versus challenging--has been demonstrated to influence inspectors' motivation and performance (Berch & Kanter 1984). Organizational effectiveness experts have long studied motivation and the incentive systems which contribute to attaining and maintaining acceptably high levels of worker motivation and performance (see, e.g., Christopher & Thor 1993 for discussions) and, as noted by Odiorne (1993):

There are certainly other motivators besides pay, and it is the favorable consequence that motivates the continuation of right behavior. Recognition, symbols of belonging, and feedback of adequacy and excellence in individual performance can be generated by appropriate supervisory practices. Money merely reinforces the general sense of approval.

Given the preceding discussion, it seems imperative that "social" influences of UT/ISI performance be included in industrial practice, efforts to upgrade requirements, methods, and techniques, and in future research. Indeed, it may be that this should be done even in preference to some of the more frequently-studied variables such as heat or noise because although the influence of these latter variables may be "somewhat reduced or modified, their existence can not be fully altered or eliminated" (Karimi 1988). That is, investments in research on social/organizational/"contextual" variables may lead to more and/or quicker improvements in ISI performance than investments in the other categories of variables detailed herein.

7 Recommendations

7.1 Recommendations Regarding Future Research

The dominant themes in this review of variables thought to impact UT/ISI performance, and as investigated in the PISC III study, may be characterized as being: equivocal results on the effects of many variables; limited applicability of laboratory findings to industrial operations; need to utilize appropriate research methodologies; and importance of management issues. There certainly is no dearth of knowledge regarding these issues, and adequate consideration of each will be required if future studies are to contribute to improving inspection performance in operational environments.

- Recommendations regarding equivocal results on the effects of many variables
 - To use Wiener's (1984) previously-cited example, future research must build on what is known in industry about avoiding time-induced performance

decrements rather than artificially creating such decrements. The results of the present review support the recommendation for a shift in emphasis away from such variables as effects of heat and noise and towards those such as managerial policies, job design, and defining UT/ISI-relevant KSAs. Similarly, the research should be more concerned with UT/ISI performance than, for example, with physiological responses.

The aforementioned work of Harris (e.g., Harris & McCloskey, 1990) which focuses on information processing abilities is an excellent example of the kind of issues which should receive increased attention. The recommendations of Wheeler, et al. (1986) and Doctor (1995; Appendix A) to base future efforts on task analytic methods are also well-founded and, if followed, will provide an important knowledge base on which to establish meaningful research programs.

- Recommendations regarding limited applicability of laboratory findings to industrial operations
 - PISC III attempted to simulate some of the relevant UT/ISI conditions in a controlled experimental environment. Such research should be continued because it may represent the best approach to gaining the knowledge necessary to improve inspection performance in the shortest time possible, i.e., as opposed to gaining knowledge regarding the basic function of environmental or other variables. It is, of course, necessary to recreate the appropriate industrial conditions (e.g., 90°F and use of protective clothing) in the applied simulation. However, the goal of improving UT/ISI performance may be better served by avoiding basic research-like explorations on variables over which industry has little control.
- Recommendations regarding the need to utilize appropriate research methodologies
 - Applied research often represents a necessary compromise between strict experimental control and full operational relevance. However, as detailed in Appendix A and as noted herein, the

PISC III research utilized a less than optimal design and in some ways did not take fullest advantage of the opportunities afforded by the design selected. Future work might best be conceived as a research program rather than as a single study. This would allow for one or more pilot studies to define appropriate levels for independent variables and to select sensitive dependent measures. It is clear from the level of sophistication in inspection tasks that the PISC III research team included individuals who were highly knowledgeable in UT/ISI practice. The addition to future research teams of one or more individuals who are equally sophisticated in research design, conduct, and analysis will enable a far greater return on research investment than was achieved by PISC III.

Overall, it is concluded that the priority should be on research which has a primary objective to develop strategies for mitigating UT/ISI performance decrements rather than to understand the mediating factors which contribute to these decrements.

- Recommendations regarding the importance of management issues
 - Variables which may be influenced by managerial structure and practice can be important contributors to improved UT/ISI performance, and applied research in these areas by organizational effectiveness experts may contribute significantly to the upgrading of UT/ISI requirements, methods, and techniques.

7.2 Recommendations Regarding Current UT/ISI Practice

- Require that such applications as certification, performance demonstration, training, and research be conducted in conditions which resemble the field as much as possible in their physical and social aspects.
- Require task analyses of UT/ISI operations and determine the KSAs required for successful inspection performance. The prioritization of these efforts should be

based upon data from Probabilistic Risk Assessments which identify the most safety-significant systems and special consideration should be given to those tasks determined through Human Reliability Analysis to be most error-prone.

- Evaluate adoption and/or modification of relevant elements of current human factors standards/guidelines from other venues (e.g., NUREG 0700; MIL-STD-1472; MIL-HDBK-759) for application to inspection. For example:
 - the recommendations regarding thermal stress reflecting consideration of ambient conditions, clothing worn, and worker activity, per NUREG 0700 "stay times"
 - the training of workers exposed to high heat in the recognition and treatment of heat illnesses, water and salt replacement, and acclimatization mechanisms, per NUREG 0700
 - the "limited tolerance" time limits of DoD MIL-STD-1472D for work in elevated temperatures
 - the recommendation in NUREG 0700 that "ear protection devices should be available and required to be worn in areas where noise levels are 85 dB or more."
- Increase the emphasis on management issues, utilizing organizational effectiveness specialists to identify and modify inappropriate philosophies, attitudes, and practices of nuclear power plant supervisors and managers.

8 References

- Bailey, R. W. *Human Performance Engineering*, 3rd Edit. Upper Saddle River, New Jersey: Prentice Hall, 1996.
- Baker, K., Olson, J., & Montgomery, J. *An Initial Examination of the Effects of Overtime and Shift Scheduling on Nuclear Power Plant Safety Performance*. Proceedings of the U.S. NRC 18th Water Reactor Safety Information Meeting, Vol. 1, 1990, pp. 57-71.
- Baker, T. L. *Alertness, Performance and Off-Duty Sleep on 8-Hour and 12-Hour Night Shifts in a Simulated Continuous Operations Control Room Setting*. NUREG/CR-6046, Washington, D.C.: U.S. Nuclear Regulatory Commission, 1996.
- Behravesh, M. M., Karimi, S. S., & Ford, M. E. *Human Factors Affecting the Performance of Inspection Personnel in Nuclear Power Plants*. Annual Review of Progress in Quantitative Nondestructive Evaluation. New York: Plenum, 1989.
- Berch, D. B., & Kanter, D. R. Individual Differences. In Warm, J. S. *Sustained Attention in Human Performance*. New York: John Wiley & Sons, 1984, pp. 143-178.
- Boff, K. R. & Lincoln, J. L. *Engineering Data Compendium, Human Perception and Performance*, Vols. 1-3. Wright-Patterson AFB, Ohio: Armstrong Laboratory, 1988.
- Christopher, W. F. and Thor, C. G. (eds.) *Handbook for Productivity Measurement and Improvement*. Portland, Oregon: Productivity Press, 1993.
- Craig, A. Human Engineering: The Control of Vigilance. In Warm, J. S. *Sustained Attention in Human Performance*. New York: John Wiley & Sons, 1984, pp. 247-292.
- Dhillon, B. S. *Human Reliability with Human Factors*. New York: Pergammon, 1986.
- Doctor, S. R. *A Critical Review of the PISC III Action 7 Report No. 31, Human Reliability in Inspection, JCN 6275-Task 1A*. Pacific Northwest National Laboratory, Dec. 29, 1995.
- Eastman Kodak Co. *Ergonomic Design for People at Work, Vol. 1*. Belmont, CA: Lifetime Learning Publications, 1983.
- Eastman Kodak Co. *Ergonomic Design for People at Work, Vol. 2*. New York: Van Nostrand Reinhold, 1986.

- Echeverria, D., Barnes, V., & Bittner, A. Jr. The Impact of Environmental Exposures on Industrial Performance of Tasks. In Karwowski, W. & Yates, J.W. (Eds.) *Advances in Industrial Ergonomics and Safety III*. Phila. PA: Taylor & Francis, 1991.
- Fleishman, E.A. & Quaintance, M.K. *Taxonomies of Human Performance*. Orlando, FL: Academic Press, 1984.
- Grandjean, E. *Fitting the Task to the Man*. Philadelphia, PA: Taylor & Francis, 1985.
- Hancock, P.A. Environmental Stressors. In Warm, J.S. *Sustained Attention in Human Performance*. New York: John Wiley & Sons, 1984, pp. 103-142.
- Harris, D.H. *Dynamic Inspection Aptitude Test User's Guide*. EPRI TR-106304, Palo Alto, CA: Electric Power Research Institute, 1996.
- Harris, D.H. *Human Performance in Nondestructive Inspections and Functional Tests*. EPRI NP-6052, Palo Alto, CA: Electric Power Research Institute, 1988.
- Harris, D.H. & McCloskey, B.P. *Cognitive Correlates of Ultrasonic Inspection Performance*. EPRI NP-6675, Palo Alto, CA: Electric Power Research Institute, 1990.
- Heasler, P.G., Taylor, T.T., Spanner, J.C., Doctor, S.R., & Deffenbaugh, J.D. *Ultrasonic Inspection Reliability for Intergranular Stress Corrosion Cracks*. NUREG/CR-4908, Washington, D.C.: U.S. Nuclear Regulatory Commission, 1990.
- Hockey, G.R. Changes in Operator Efficiency as a Function of Environmental Stress, Fatigue, and Circadian Rhythms. In Boff, K.R., Kaufman, L., & Thomas, J.P. *Handbook of Perception and Human Performance, Vol. 2, Cognitive Processes and Performance*. New York: John Wiley & Sons, 1986, pp. 44.1-44.49.
- Karimi, S. S. *Human Factors Affecting the Performance of Inspection Personnel in Nuclear Power Plants*. EPRI P-6000, Palo Alto, California: Electric Power Research Institute, 1988.
- Latack J. C. & Foster, L. W. Implementation of Compressed Work Schedules: Participation and Job Redesign as Critical Factors for Employee Acceptance. *Personnel Psychology*, 1985, 38, 75-91.
- Luczak, H. & Mueller, T. Worker Attention and Fatigue. In Salvendy, G. & Karwowski, W. *Design of Work and Development of Personnel in Advanced Manufacturing*. New York: John Wiley & Sons, 1994, pp. 463-491.
- Mahan, R. P. Continuous Performance and Operator Judgement Implications for the Automated Work Place. In Karwowski, W. & Rahimi, M. *Ergonomics of Hybrid Automated Systems II*. Amsterdam: Elsevier, 1990, pp. 741-746.
- Megaw, E. D. *Some Factors Affecting Inspection Accuracy*. Presentation at Symposium on Ergonomics and Visual Inspection. Birmingham, England, 1978.
- Monden, Y. Characteristics of Performance Control Systems in Japanese Corporations. In Christopher, W. F. and Thor, C. G. (eds.) *Handbook for Productivity Measurement and Improvement*. Portland, Oregon: Productivity Press, 1993, pp. 8.9.1-8.9.8.
- Moray, N. *The Role of Attention in the Detection of Errors and the Diagnosis of Failures in Man-machine Systems*. In Rasmussen, J. & Rouse, W. B. *Human Detection and Diagnosis of System Failures*. New York: Plenum, 1991, pp. 185-198.
- Murray, R. H., & McCally, M. Combined Environmental Stressors. In Parker, J. F. Jr., & West, V. R. *Bioastronautics Data Book*. Washington, D.C.: National Aeronautics and Space Administration, 1973, pp. 881-914.
- Odiorne, G. S. Eleven Steps to Making a New Management Program Succeed. In Christopher, W. F. and Thor, C. G. (eds.) *Handbook for Productivity Measurement and Improvement*. Portland, OR: Productivity Press, 1993, pp. 2.5.1-2.5.6.
- Parasuraman, R. Vigilance, Monitoring, and Search. In Boff, K. R., Kaufman, L., & Thomas, J. P. *Handbook of Perception and Human Performance, Vol. 2, Cognitive Processes and Performance*. New York: John Wiley & Sons, 1986, pp. 45.1-44.35.

- Parasuraman, R. The Psychobiology of Sustained Attention. In Warm, J. S. *Sustained Attention in Human Performance*. New York: John Wiley & Sons, 1984, pp. 61-102.
- Parasuraman, R. Vigilance, Arousal, and the Brain. In Gale, A. & Edwards, J. A. *Physiological Correlates of Human Behavior, Vol. 2, Attention and Performance*. London: Academic Press, 1983, 35-56.
- Park, K. S. Human Error. In Salvendy, G. (ed.), *Handbook of Human Factors and Ergonomics, 2nd Edit.* New York: Wiley, 1997.
- Pierce, J. L. & Dunham, R. B. *The 12-Hour work Day: A 48-Hour, Eight-Day Week*. Academy of Management Journal 1992, 35(5), 1086-1098.
- Pierce, J. L., Newstrom, J. W., Dunham, R. B., & Barber, A. E. *Alternative Work Schedules*. Boston: Allyn & Bacon, 1989.
- PISC (Programme for the Inspection of Steel Components). *Human Reliability in Inspection, Final report on Action 7 in the PISC III Programme*, PISC III Report 31. Organization for Economic Co-operation and Development, Nuclear Energy Agency, Committee on the Safety of Nuclear Installations, 1994.
- Proctor, R. W. & Van Zandt, T. *Human Factors in Simple and Complex System*. Boston: Allyn & Bacon, 1994.
- Rasmussen, J. Duncan, K., & Leplat, J. *New Technology and Human Error*. New York: Wiley & Sons, 1987.
- Sanders, M.S. & McCormick, E.J. *Human Factors in Engineering and Design*. New York: McGraw Hill, 1987.
- Spanner, J. C., Badalamente, R. V., Rankin, W. L., & Triggs, T. J. *Human Reliability Impact on Inservice Inspection*. NUREG/CR-4436 Vol. 1, Washington, D.C.: U.S. Nuclear Regulatory Commission, 1986.
- Tepas, D. I. Flextime, Compressed Workweeks and Other Alternative Work Schedules. In Folkard, S. & Monk, T. H. *Hours of Work*. New York: John Wiley & Sons, 1985 pp. 147-164.
- Tepas, D. I. & Monk, T. G. Work Schedules. In Salvendy, G. *Handbook of Human Factors*. New York: John Wiley & Sons, 1987, pp. 820-843.
- Triggs, T. J., Rankin, W. L., Badalamente, R. V., & Spanner, J. C. *Human Reliability Impact on Inservice Inspection*. NUREG/CR-4436 Vol. 2, Washington, D.C.: U.S. Nuclear Regulatory Commission, 1986.
- U.S. Department of Defense. *Human Factors Engineering Design Guide for Army Material, MIL-HDBK-59A*. Washington, D.C., U.S. Government Printing Office, 1989.
- U.S. Department of Defense. *Human Engineering Design Criteria for Military Systems, Equipment, and Facilities, MIL-STD-1472D*. Washington, D.C., U.S. Government Printing Office, 1989.
- U.S. Nuclear Regulatory Commission. *Human-System Interface Design Review Guideline, Process and Guidelines*. NUREG-0700 Rev.1. Vol. 1, Washington, D.C.: U.S. Nuclear Regulatory Commission, 1996.
- Warm, J. S. An Introduction to Vigilance. In Warm, J. S. *Sustained Attention in Human Performance*. New York: John Wiley & Sons, 1984, pp. 1-14.
- Webster, C. Some Individual Psychological Factors Leading to Error in Ultrasonic Testing. *British Journal of Non-destructive Testing*, 1989, 31(12), 680-82.
- Wheeler, W. A., Rankin, W. L., Spanner J. C., Badalamente, R. V., & Taylor, T. T. *Human Factors Study Conducted in Conjunction with a Mini-Round Robin Assessment of Ultrasonic Technician Performance*, NUREG/CR-4600, Washington, D.C.: U.S. Nuclear Regulatory Commission, 1986.
- Wiener, E. L. Vigilance and Inspection. In Warm, J. S. *Sustained Attention in Human Performance*. New York: John Wiley & Sons, 1984, pp. 207-246.
- Woodson, W. E., Tillman, B., & Tillman, P. *Human Factors Design Handbook. 2nd Edit.* New York: McGraw Hill, 1992.

A Critical Review of the PISC III Report No. 31 Human Reliability in Inspection

INTRODUCTION

Research and inspection experience has shown that the ability of inspectors to reliably detect all flaws important to structural integrity while performing nondestructive examination (NDE) during inservice inspection is less than ideal. Further, data from the Programme for the Inspection of Steel Components Phase 2 (PISC II) suggested that there may be large differences in the skills, knowledge, abilities, and working practices of personnel performing manual NDE and these could significantly affect the reliability of the inspection. As a consequence of these factors, a study investigating human reliability issues was initiated as Action 7 of the PISC III Program. The objectives of this research were to identify causes of observed variability in inspection performance and to identify factors influencing the reliability of inspections in nuclear power plants. The purpose of this report is to document a review of the PISC III report "Human Reliability in Inspection".

This report will first evaluate the research methods employed to investigate the factors that may contribute to human reliability in NDE and then will comment on the conclusions that were drawn from the research data. Suggestions for improvements in the design of research for NDE inspections will be offered.

PISC III STUDY DESIGN

Research Philosophy

The underlying philosophy of the PISC III research team was to design an experiment that would provide tasks that were representative of inservice inspection in nuclear power plants. The basic approach to the experiment was to compare inspection performance in a benign environment to that in a simulated nuclear power plant environment. It was thought that differences in performance could then be attributed to either differences in the task conditions or changes in an inspector's physiological or psychological condition.

To this end, the PISC III researchers took the following steps in designing the study:

- professionally qualified inspectors were employed as subjects
- test plates with a number of planar flaws were fabricated for inspection purposes
- a computer-based simulator was employed to permit the study of a range of inspection skills
- well-accepted procedures for NDE inspection were utilized
- a work environment with some of the attributes of a nuclear power plant was created.

Subjects

The subjects for PISC III study were selected as a part of a PISC Operating Agent solicitation to inspection organizations in Europe. All subjects were professional inspectors qualified at a minimum of an American Society for Nondestructive Testing (ASNT) Level 2 or a national equivalent. Six subjects were selected from the solicitation to PISC member nations and their participation in the research study was paid. These inspectors' training and experience were representative of NDE inspectors in the nuclear industry. Selection of professional inspectors for subjects in the experiment served the research well in that it ensured that the subjects were qualified to perform the procedures that would be required during the experiment.

Research Facilities

The study was conducted at research facilities at Risley, UK, where two inspection environments were created to investigate factors affecting human reliability in NDE inspection tasks. The first environment was a laboratory environment or Studio set up to provide to the extent possible optimal conditions for the inspectors to perform the experimental tasks. The Studio was a quiet room where the temperature was controlled between 18-23°C (64-73°F) and humidity was that of the ambient environment. Subjects were required to perform the experimental inspection tasks during an eight-hour day for five consecutive days. Breaks for lunch and in the morning and afternoon were given to the subjects. The Studio was a comfortable environment for the inspectors to perform their experimental inspection tasks. In addition to providing a good task environment, NDE inspection scans were performed with the test plates in a horizontal orientation which is considered to be the easiest of the conditions found in an inservice inspection.

The second experimental facility was set up at Risley in a Transportable Environmental Laboratory (TEL) to simulate conditions that are more representative of a realistic inspection environment. The goal was to increase the environmental stressors on the inspectors to those more representative of inspections performed in nuclear power plants to determine if these significantly affected the reliability of their inspection performance. To this end, the researchers:

- increased the temperature of the test facility to 25-27°C (77-81°F),
- raised the humidity to 40-50%,
- reduced the illumination level, and
- provided a background noise level of 80 dBA by playing recordings of noise from a nuclear power plant where inservice inspection is normally performed.

Inservice inspection conditions were further simulated by lengthening the inspection day to 12 hours and increasing the number of inspection days to 12 with a day off following the sixth day of the experimental trials in the TEL. Task complexity was also increased in the TEL by requiring the inspectors to scan with the test plates oriented vertically rather than horizontally.

The variations in the task environment demonstrate that the researcher had a basic understanding of the need to establish a baseline for inspection performance prior to attempting to estimate the effect of the environmental stressors in a nuclear environment. However, there are two concerns with this aspect of the experimental design. The first is that while lengthening the work day could be expected to affect performance, the increases in temperature, humidity, and noise level only marginally changed the environmental conditions. The NDE inspection task is not particularly a physically demanding task, but rather one that stresses an inspector's perceptual and cognitive abilities. As a result, large differences in task performance due to the modifications in the environmental conditions would not have been predicted by most scientific literature. More attention could have been paid to the relationship between the task requirements and the expected performance decrement due to environmental conditions prior to setting the parameters for the test conditions. A second concern is that both the environment and the task complexity changed when the inspectors performed their tasks in the TEL. The experimental design confounded these factors in a way that makes it impossible to infer whether task complexity or the more severe environment contributed to observed differences in performance in the TEL when compared to performance in the Studio. Separate manipulation of these variables would have significantly added to the interpretability of the study data:

Experimental Tasks

The experimental tasks that were used in the study are those that would typically be performed by NDE inspectors conducting an inservice inspection. Test plates and a simulation were constructed to permit the subjects to perform the procedures that normally guide an NDE inspection. To add to the realism of the simulation, the inspection procedures followed during the experiment were modeled after accepted inspection practices.

NDE inservice inspection has several task components and each was represented in the experimental tasks that were performed by the subjects. The task components included in the design were: calibration, initial scanning, maximizing initial flaw indications and comparing them with recording criteria, flaw sizing, plotting, and reporting. The methods used to study each of the task components and the measures used to evaluate inspector performance in these areas are discussed in the following paragraphs.

Calibration was investigated in two ways. A pre-inspection calibration activity was performed both prior to and after scanning a test plate. A second experimental task required the inspector to perform a Caltest, a series of 15 calibrations using standard calibration equipment. Deviation from the expected value, as measured in dB, was used as the dependent measure to evaluate calibration performance.

Scanning and signal optimization were also studied using two tasks. In the first test, inspectors were asked to scan test plates using the provided NDE equipment. The inspectors followed accepted inspection procedures during the test. Initial scanning was used to find any flaws in the test plate. Potential flaws were marked during the initial scan and returned to after completing the

scan. The signals from the flaws were optimized according to procedure and compared to reference data. Data from this portion of the experiment were reduced to a flaw detection frequency, the percentage of flaws detected when compared to the total number of flaws in the test plate. A second scanning test utilized a computer-based inspection simulator. Again the inspectors were asked to scan a weld and report the presence and magnitude of the flaws detected. In addition to collecting flaw detection frequency data for the simulator trials, scanning pattern and sensor coupling were also evaluated.

Plotting was studied at the end of each day's experimental session. Each inspector was asked to plot the results of his or her daily scans to determine the size and location of the detected flaws. The plotting task was subdivided into six subtasks. An error in one of the six tasks was reported if mistakes were made in the plotting task.

The reporting component of the inspection task was investigated by examining the data sheets that were produced by the inspectors during the course of the day's study. These were examined by researchers to identify places where further examination of study data such as video records or calibration records was needed to determine the causes of a non-detection.

The use of accepted inspection procedures in the performance of the experimental inspections and the evaluation of task components demonstrates that the researchers in this study had a very good understanding of the characteristics that were required to provide a realistic task environment. Attempts were additionally made to divide the major task components of an NDE inservice inspection to ensure that data were collected that would describe inspection performance for each of the task components.

The measures selected to describe the performance in these tasks, however, make it difficult to identify the reasons performance varied from condition to condition, day to day, or inspector to inspector. Flaw detection frequency, the principal measure used in the study, was an outcome-based measure in that it indicated overall performance in the task, rather than assessing the components of the behavior that resulted in a flaw detection or a miss. It quantified the overall impact to the inspection, but provided no insight into how task performance might be improved. Other measures used in the study had similar characteristics. The identification of causes of variability in inspection performance or the causes of error requires the selection of measures better focused at the inspection task components and the inspector's task requirements. For example, a measure that related a missed flaw to the loss of coupling in the scan would have yielded data much more amenable to cause identification than flaw detection frequency. While the latter is an important measure in an operational environment, it contributed little to achieving the study's goals.

Psychological and Physiological Data

In addition to these performance data, psychological tests were administered to assess mental outlook, attitude, and aptitude for the inspection task and physiological data were collected to evaluate the inspector's level of tiredness during the course of the experiment.

The psychological tests that were conducted as a part of this study included questionnaire data at the beginning and end of each day's testing, an arithmetic test, a spatial visualization test, the Eyesenk Personality Inventory, a general condition questionnaire, an aptitude test, and a visual acuity test. Data were used to attempt to explain inspector performance during the course of the experiment.

Electrocardiogram, pulse rate, and deep body temperature were collected for some subjects in the experiment. These physiological measures were collected in an attempt to determine a level of "tiredness" of the individual and measure inspector fitness as the test proceeded. The research team hoped that the environmental stressors that were induced in the TEL would be reflected in these measures. Unfortunately, no changes were observed early in the testing and the data collection was discontinued during the course of the study.

The overall concept of performance decrements being reflected in physiological measures was a good one. However, the expectation that these physiological variables should change as a result of the conditions present in the TEL when compared to the Studio is unfounded in the literature. The problems associated with this part of the experiment, however, are that the choice of measures was apparently not based on research which showed them to be sensitive to these sorts of changes and that these data were not collected during the entire study or for all inspectors. Both the collection of physiological data for only a subset of the inspectors and discontinuing the collection of physiological data midway through the experiment changed the conditions of the experiment and may have contributed to any observed performance differences. The collection of physiological data is obtrusive and in the case of deep body temperature intrusive and it can be expected to change the way the subject performs the task.

Experimental Controls

Control over the experimental trials was governed by a quality control procedure and data were collected by test administrators, called invigilators. The invigilators' duties were guided by a procedure and included monitoring of the test program, providing written comments on the capability and performance of the inspector, monitoring the medical equipment, and controlling the supply and receipt of inspection procedures and results sheets. The invigilators possessed and maintained the inspection tasks that each subject was to complete each day and ensured that the inspection program was followed. A one-way glass was used to observe the inspectors, and multiple video records of the trials were taken for later analysis, thereby reducing the effect of the experimental environment on inspector performance.

While the report documents the controls to ensure that the data were properly collected and reported, it is curious that the number of trials or inspection periods is not equal across the subjects in the experiment. The number of calibration tests and the number of simulator trials also indicate that inspectors were permitted to omit part of the research program during the course of the experiment. These missing data, together with a small sample size, make it difficult to draw conclusions from the study.

STUDY RESULTS

Study results were reported by task component in the PISC III report. Conclusions drawn from the study data will be organized similarly in this report.

Calibration. Conclusions drawn from the calibration data include the following:

- errors committed in calibrating the NDE equipment can be separated into measurement error and human error
- measurement error was not affected by the environmental conditions presented in the TEL and is acceptably small
- human error was greater in the TEL and was influenced by the tiredness of the inspector.

A brief re-analysis of the data contained in the report tends to support the conclusion that measurement error in the calibration task was not affected by the conditions presented in the TEL. It did not support the conclusion that human error increased as a result of conditions in the TEL or that inspector fatigue contributed to increased error rates. In our analysis means and standard deviations were calculated for the Caltest trials (Table 1). While the data were incomplete, performance as indicated by these measures was similar for the Studio and the TEL. Further, performance in the morning and the afternoon in the TEL was almost identical which suggests that inspector tiredness was not a factor in the calibration tests. It appears the conclusions drawn for this portion of the experiment were based on a "scan" of the data rather than statistical analysis and that the authors emphasized the calibration data of one or two inspectors in drawing their conclusion with respect to human error and inspector tiredness. This is particularly problematic with the small sample size used in this study.

	Morning		Afternoon	
	Average	Standard Dev	Average	Standard Dev
Week 1	0.60	0.15	0.73	0.22
Week 2	0.60	0.17	0.65	0.13
Overall	0.60	0.15	0.70	0.19

Table 1 - Caltest Data - TEL

Plate Inspections. Conclusions from the plate inspection data include:

- there were no major differences in performance in the Studio and the TEL
- inspectors experienced an initial period of adjustment
- detection rate falls midway through the second week and this is due to tiredness
- detection performance is related to the magnitude of the flaw in the weld.
- there is no correlation between work rate and performance.

In order to evaluate the plate inspection conclusions, plate inspection data contained in Appendix A were reanalyzed. Means and standard deviations for flaw detection frequency were calculated for trials performed in the Studio (Table 2) and the TEL (Table 3). Throughput data for number of plate tests performed (Table 4) were also analyzed. These analyses support the conclusion that there were no major differences in the inspection performance in the TEL when compared to the Studio. No support was shown for an initial adjustment period, a reduction in inspection performance during the second week, or for tiredness contributing to a reduction in inspection performance. When the mean flaw detection frequency and the standard deviation for flaw detection frequency in Week One are compared to Week Two and Morning performance using these statistics is compared to Afternoon performance, there appear to be no major differences in inspection performance. The means and standard deviations for the plate inspection are approximately equal. The overall variance among the subjects is sufficiently high that even though the plotted data indicate differences in performance, all data are within the experimental error.

It is somewhat surprising that the performance in the first week did not differ from performance in the second week and that differences in performance from the Studio and the TEL were not found. While the changes in environmental conditions may not have contributed to performance differences, one would have predicted a performance difference due to the change in task complexity. The ease of the task in the Studio is indicated by the 100% detection performance achieved by many of the inspectors. The fact that performance differences were not observed may indicate that scanning in the vertical orientation was still a comparatively easy task when compared with their normal inspection environment or that the Studio provided a high transfer of training effect for the tasks in the TEL. A second factor that may have masked differences in performance in the two environments is variability in performance due to collecting physiological data for only a portion of the study. The increased variance may have masked small real changes in performance levels. Changes in the experimental design would have been necessary to isolate either of these effects.

	Average	Standard Dev
Studio	89.5	15.1

Table 2 - Flaw Detection Frequency - Studio

	Average	Standard Dev
Week 1	85.8	13.8
WeeK 2	87.0	15.2
Overall	86.3	14.4

Table 3 - Flaw Detection Frequency - TEL

	Average	Standard Dev
Week 1	8.8	3.9
WeeK 2	9.9	4.7
Overall	9.5	4.3

Table 4 -Throughput (plate tests) - TEL

Simulator Tests. Conclusions from the simulator data include:

- flaws are not detected for a variety of reasons including improper scan overlap, poor coupling, the signal being missed by the inspector, or the inspector looking away from the screen when the signal was displayed
- tiredness contributed to deterioration in inspection performance
- the probability of detection is correlated to the amplitude of the signal
- additional sensitivity would improve detection, especially for small flaws.

The simulator data were reanalyzed in the same manner as the plate inspection data by calculating means and standard deviations across the six inspectors. These analyses indicate that there are no major differences in inspection performance when Week One data is compared to Week Two and

when performance in the Afternoon is compared to the Morning (see Table 5). It appears the researchers again drew conclusions from scanning the plots of the data and that too much emphasis was placed on individual performance and isolated data points when interpreting the data. This data analysis strategy led to inferring effects that can not be supported by the overall data collected in the study.

However, several interesting observations can be made from the simulator data. The simulator tests demonstrated that even experienced inspectors have a great deal of difficulty maintaining the proper scan paths and coupling and that these two factors led to a reduction in detection rate. Because experienced inspectors were used in the study, it is unlikely that these difficulties result from deficiencies in inspection skills. It is possible that the task requirements involved in maintaining the scan path and coupling exceed human abilities to reliably perform the task. Changes in inspection methods or equipment may be necessary to resolve the observed mismatch.

Additionally, PISC review of the video tapes indicate that many signals indicating flaws are both brief and weak and, as a consequence, are missed by inspectors. These findings also have implications for the design of inspection equipment and training. Design modifications to guide or verify the scan path, the ability to store the flaw signal for later review, or aids to assist in marking the flaws may significantly improve the inspectors' ability to locate small flaws. The simulator experience further suggests that simulators could be used as an effective training tool for NDE inspectors. Feedback on the efficacy of the scan during the simulator trials may have significantly improved inspection performance in the subsequent trials.

Lastly, an opportunity to compare the differences between good scanning practices and poorer ones was lost in this study. The video recordings of inspectors that performed well in the simulator trials could have been compared to those who performed poorly. Differences in scanning behavior could have been noted and perhaps been used to suggest new procedures and training to improve an inspector's ability to reliably perform inservice inspections.

	Morning		Afternoon	
	Average	Standard Dev	Average	Standard Dev
Week 1	0.66	0.24	0.70	0.24
Week 2	0.81	0.25	0.84	0.21
Overall	0.73	0.26	0.75	0.24

Table 5 -Flaw Detection Frequency - Simulator

Plotting Test Results. The following were conclusions from the plotting tests:

- a major source of mistakes in plotting and reporting was in the drawing of the beam angle
- drawing aids should be provided to the inspectors to reduce the likelihood of plotting errors.

The data collected in this portion of the study were particularly useful because the plotting task was divided into its major components. Errors in these components were then used as measures of performance in the plotting task. By evaluating inspector performance at a task component level rather than at the whole task level, researchers were able to suggest more concrete solutions to improve inspector performance for this part of the NDE inspection task.

Psychological Tests. The following were conclusions from the psychological tests:

- the daily questionnaires provided useful information concerning the subjects' state of mind and physical condition
- the de-briefing questionnaire indicated that all of the inspectors became tired during a shift
- the length of work shift and repetitiveness of the work contribute to the feeling of tiredness.

Because the selection of the psychological assessments was not done in a manner that specifically related task requirements to the behavior being assessed, any conclusions drawn from these data are highly speculative. The small sample size and the large effect that any one subject can have on the interpretation of the data casts even further doubt on any interpretation. It appears that the attribution of tiredness leading to inspection performance decrement is taken solely from self-reports in a debriefing questionnaire that was administered each day. Since the performance data do not reflect a significant performance decrement and the method for determining inspector tiredness was not reported, these assertions should not be made.

Trying to relate personality and cognitive variables to inspector performance through the use of questionnaire data and psychological testing is a reasonable test strategy. It appears, however, that the tests selected and data collected had no substantiated relationship to the tasks involved in NDE inspection, but rather were a product of the notion that if the data were collected, it might be useful in explaining some aspects of inspector performance, or it would quantify the magnitude of variability due to the human aspects.

The research team might have been better served performing a job analysis, specifying the knowledge, skills, and abilities needed to successfully perform inspections. This would have permitted them to infer the type of spatial, math or other abilities and the personality attributes of individuals that are likely to be successful inspectors. From these data, appropriate tests could be selected and administered. These data could then be correlated with the job performance

dimensions that are considered to be critical to inspector performance and contribute to the development of a better understanding of the relationships among the many factors affecting inspector performance.

SUMMARY

The PISC researchers put forth a large effort into investigating factors that might affect inspector performance in inservice inspection. The study recreated a nuclear power plant inspection environment in which experienced and qualified NDE inspectors were permitted to conduct inspections according to accepted procedures. The inspection environment and inspection complexity were varied to investigate their effects on inspection performance as indicated by flaw detection frequency. While the researchers understood the NDE inspection task and some of the rudimentary considerations of the design of human factors experiments, the study contained many design flaws that made it impossible to isolate the effect that inspection environment, the main factor in this study, had on observed inspection performance. Some of the problems with the experimental design include:

- changes in the task environment due to discontinuing physiological data collection
- physiological data not being collected for all subjects,
- incomplete experimental trials for some subjects which led to empty data cells
- confounding task complexity with environmental conditions in the design.

Lastly, all too often conclusions in this study were based on individual results or subject data that tended to differ significantly from the mean. A better understanding of the design of behavioral experiments could have prevented the confounds in data collection, improved the interpretability of the data in the study, and furthered the objectives and goals stated for the study.

Overall, not much can be concluded from the data collected in the PISC study. The data supported no differences in performance in the TEL when compared to the Studio environment, performance in the first week was roughly the same as performance in the second week, performance in the morning was similar to that in the afternoon. Data from other task environments suggest that the environment, tiredness, task complexity, and a host of other human factors can affect performance and the reliability of performance in a nuclear power plant setting. Unfortunately, this study did little to increase our understanding of these factors in the inservice inspection tasks.

Findings from the simulator trials indicate that maintaining scan path and coupling is difficult and inspectors tend to miss short duration and weak signals. These data suggest both training and design interventions are necessary. The research also suggests that performance in the plotting and reporting functions can benefit from job performance aids. Lastly, it can be concluded that the inspectors reported that the long hours and the repetitive nature of the inspection tasks in the experiment made them tired at the end of the day. However, the extent to which physical, emotional, cognitive, or perceptual factors contributed to this tiredness is unclear from this study.

NDE inspection reliability is an important issue in the nuclear industry and this research demonstrates that a realistic inspection environment can be created. If we are to better understand the factors that lead to errors in inservice inspection, studies like this one, but utilizing a more suitable experimental design, will be necessary.

Future work should be built on lessons learned from behavioral research concerned with inspection in work environments like those found in the process and power industries. Review of these data will provide the information necessary to define the independent variables that most contribute to inspector variability and assist in setting appropriate levels for each of the factors. To augment prior research, a survey of inspectors to identify factors they believe to be impactful should be conducted. Also, observation of inservice inspection either in real time or via video tape to identify the difficulties that inspectors are experiencing may be beneficial. Finally, future work should seek to use an appropriate experimental design to isolate the variables that have been identified as impacting inspection reliability, and studies must employ sufficient numbers of subjects so that the results can be interpreted in a manner meaningful to the nuclear power plant inservice inspection environment.

Appendix B

Updated Definitions for the Ability Categories in Recent Forms of the Manual for the Ability Requirements Scale (MARS)¹

1. *Oral Comprehension* is the ability to understand spoken English words and sentences.
2. *Written Comprehension* is the ability to understand written sentences and paragraphs.
3. *Oral Expression* is the ability to use English words or sentences in speaking so others will understand.
4. *Written Expression* is the ability to use English words or sentences in writing so others will understand.
5. *Fluency of Ideas* is the ability to produce a number of ideas about a given topic.
6. *Originality* is the ability to produce unusual or clever ideas about a given topic or situation. It is the ability to invent creative solutions to problems or to develop new procedures to situations in which standard operating procedures do not apply.
7. *Memorization* is the ability to remember information, such as words, numbers, pictures, and procedures. Pieces of information can be remembered by themselves or with other pieces of information.
8. *Problem Sensitivity* is the ability to tell when something is wrong or is likely to go wrong. It includes being able to identify the whole problem as well as the elements of the problem.
9. *Mathematical Reasoning* is the ability to understand and organize a problem and then to select a mathematical method or formula to solve the problem. It encompasses reasoning through mathematical problems to determine appropriate operations that can be performed to solve problems. It also includes the understanding or structuring of mathematical problems. The actual manipulation of numbers is not included in this ability.
10. *Number Facility* involves the degree to which adding, subtracting, multiplying, and dividing can be done quickly and correctly. These can be steps in other operations like finding percentages and taking square roots.
11. *Deductive Reasoning* is the ability to apply general rules to specific problems to come up with logical answers. It involves deciding if an answer makes sense.
12. *Inductive Reasoning* is the ability to combine separate pieces of information, or specific answers to problems, to form general rules or conclusions. It involves the ability to think of possible reasons for why things go together.
13. *Information Ordering* is the ability to follow correctly a rule or set of rules to arrange things or actions in a certain order. The rule or set of rules used must be given. The things or actions to be put in order can include numbers, letters, words, pictures, procedures, sentences, and mathematical or logical operations.

¹ From Fleishman & Quaintance (1984).

14. *Category Flexibility* is the ability to produce many rules so that each rule tells how to group a set of things in a different way. Each different group must contain at least two things from the original set of things.

15. *Speed of Closure* involves the degree to which different pieces of information can be combined and organized into one meaningful pattern quickly. It is not known beforehand what the pattern will be. The material may be visual or auditory.

16. *Flexibility of Closure* is the ability to identify or detect a known pattern (like a figure, word, or object) that is hidden in other material. The task is to pick out the disguised pattern from the background material.

17. *Spatial Orientation* is the ability to tell where you are in relation to the location of some object or to tell where the object is in relation to you.

18. *Visualization* is the ability to imagine how something will look when it is moved around or when its parts are moved or rearranged. It requires the forming of mental images of how patterns or objects would look after certain changes, such as unfolding or rotation. One has to predict how an object, set of objects, or pattern will appear after the changes are carried out.

19. *Perceptual Speed* involves the degree to which one can compare letters, numbers, objects, pictures, or patterns, quickly and accurately. The things to be compared may be presented at the same time or one after the other. This ability also includes comparing a presented object with a remembered object.

20. *Control Precision* is the ability to move controls of a machine or vehicle. This involves the degree to which these controls can be moved quickly and repeatedly to exact positions.

21. *Multilimb Coordination* is the ability to coordinate movements of two or more limbs (for example, two arms, two legs, or one leg and one arm), such as in moving equipment controls. Two or more limbs are in motion while the individual is sitting, standing, or lying down.

22. *Response Orientation* is the ability to choose between two or more movements quickly and accurately when two or more different signals (lights, sounds, pictures) are given. The ability is concerned with the speed with which the right response can be started with the hand, foot, or other parts of the body.

23. *Rate Control* is the ability to adjust an equipment control in response to changes in the speed and/or directions of a continuously moving object or scene. The ability involves timing these adjustments in anticipating these changes. This ability does not extend to situations in which both the speed and direction of the object are perfectly predictable.

24. *Reaction Time* is the ability to give one fast response to one signal (sound, light, picture) when it appears. This ability is concerned with the speed with which the movement can be started with the hand, foot, or other parts of the body.

25. *Arm-Hand Steadiness* is the ability to keep the hand and arm steady. It includes steadiness while making an arm movement as well as while holding the arm and hand in one position. This ability does not involve strength or speed.

26. *Manual Dexterity* is the ability to make skillful coordinated movements of one hand, a hand together with its arm, or two hands to grasp, place, move, or assemble objects like hand tools or blocks. This ability involves the degree to which these arm-hand movements can be carried out quickly. It does not involve moving machine or equipment controls like levers.

27. *Finger Dexterity* is the ability to make skillful, coordinated movements of the fingers of one or both hands and to grasp, place, or move small objects. This ability involves the degree to which these finger movements can be carried out quickly.

28. *Wrist-Finger Speed* is the ability to make fast, simple repeated movements of the fingers, hands, and wrists. It involves little, if any, accuracy or eye-hand coordination.

29. *Speed of Limb Movement* involves the speed with which a single movement of the arms or legs can be made. This ability does not include accuracy, careful control, or coordination of movement.

30. *Selective Attention* is the ability to concentrate on a task one is doing. This ability involves concentrating while performing a boring task and not being distracted.

31. *Time Sharing* is the ability to shift back and forth between two or more sources of information.

32. *Static Strength* is the ability to use muscle force in order to lift, push, pull, or carry objects. It is the maximum force that one can exert for a brief period of time.

33. *Explosive Strength* is the ability to use short bursts of muscle force to propel oneself or an object. It requires gathering energy for bursts of muscle effort over a very short time period.

34. *Dynamic Strength* is the ability of the muscles to exert force repeatedly or continuously over a long time period. This is the ability to support, hold up, or move the body's own weight and/or objects repeatedly over time. It represents muscular endurance and emphasizes the resistance of the muscles to fatigue.

35. *Trunk Strength* involves the degree to which one's stomach and lower back muscles can support part of the body repeatedly or continuously over time. The ability involves the degree to which these trunk muscles do not fatigue when they are put under such repeated or continuous strain.

36. *Extent Flexibility* is the ability to bend, stretch, twist, or reach out with the body, arms, or legs.

37. *Dynamic Flexibility* is the ability to bend, stretch, twist, or reach out with the body, arms and/or legs, both quickly and repeatedly.

38. *Gross Body Coordination* is the ability to coordinate the movement of the arms, legs, and torso together in activities in which the whole body is in motion.

39. *Gross Body Equilibrium* is the ability to keep or regain one's body balance or to stay upright when in an unstable position. This ability includes maintaining one's balance when changing direction while moving or standing motionless.

40. *Stamina* is the ability of the lungs and circulatory systems of the body to perform efficiently over long time periods. This is the ability to exert oneself physically without getting out of breath.

41. *Near Vision* is the capacity to see close environmental surroundings.

42. *Far Vision* is the capacity to see distant environmental surroundings.

43. *Visual Color Discrimination* is the capacity to match or discriminate between colors. This capacity also includes detecting differences in color purity (saturation) and brightness (brilliance).

44. *Night Vision* is the ability to see under low light conditions.

45. *Peripheral Vision* is the ability to perceive objects or movement towards the edges of the visual field.

46. *Depth Perception* is the ability to distinguish which of several objects is more distant from or nearer to the observer, or to judge the distance of an object from the observer.

47. *Glare Sensitivity* is the ability to see objects in the presence of glare or bright ambient lighting.

48. *General Hearing* is the ability to detect and to discriminate among sounds that vary over broad ranges of pitch and/or loudness.

49. *Auditory Attention* is the ability to focus on a single source of auditory information in the presence of other distracting and irrelevant auditory stimuli.

50. *Sound Localization* is the ability to identify the direction from which an auditory stimulus originated relative to the observer.

51. *Speech Hearing* is the ability to learn and understand the speech of another person.

52. *Speech Clarity* is the ability to communicate orally in a clear fashion understandable to a listener.

BIBLIOGRAPHIC DATA SHEET

(See instructions on the reverse)

1. REPORT NUMBER
(Assigned by NRC, Add Vol., Supp., Rev.,
and Addendum Numbers, if any.)

NUREG/CR-6605
PNNL-11797

2. TITLE AND SUBTITLE

An Evaluation of Human factors Research for
Ultrasonic Inservice Inspection

3. DATE REPORT PUBLISHED

MONTH	YEAR
March	1998

4. FIN OR GRANT NUMBER

W6275

5. AUTHOR(S)

D.J. Pond, D.T. Donohoo, R.V. Harris, Jr.

6. TYPE OF REPORT

Technical

7. PERIOD COVERED (Inclusive Dates)

8. PERFORMING ORGANIZATION - NAME AND ADDRESS (If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)

Pacific Northwest National Laboratory
Richland, WA 99352

9. SPONSORING ORGANIZATION - NAME AND ADDRESS (If NRC, type "Same as above"; if contractor, provide NRC Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address.)

Division of Engineering Technology
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

10. SUPPLEMENTARY NOTES

D. Jackson, NRC Project Manager

11. ABSTRACT (200 words or less)

This work was undertaken to determine if human factors research has yielded information applicable to upgrading requirements in ASME Boiler and Pressure Vessel Code Section XI, improving methods and techniques in Section V, and/or suggesting relevant research. A preference was established for information and recommendations which have become accepted and standard practice. Manual Ultrasonic Testing/Inservice Inspection (UT/ISI) is a complex task subject to influence by dozens of variables. This review frequently revealed equivocal findings regarding effects of environmental variables as well as repeated indications that inspection performance may be more, and more reliability, influenced by the workers' social environment, including managerial practices, than by other situational variables. Also of significance are each inspector's relevant knowledge, skills, and abilities, and determination of these is seen as a necessary first step in upgrading requirements, methods, and techniques as well as in focusing research in support of such programs. While understanding the effects and mediating mechanisms of the variables impacting inspection performance is a worthwhile pursuit for researchers, initial improvements in industrial UT/ISI performance may be achieved by implementing practices already known to mitigate the effects of potentially adverse conditions.

12. KEY WORDS/DESCRIPTORS (List words or phrases that will assist researchers in locating the report.)

Ultrasonic Testing, Inservice Inspection, Inspection Performance, Task Analysis,
Abilities Assessment, Human Factors, Organizational Effectiveness

13. AVAILABILITY STATEMENT

unlimited

14. SECURITY CLASSIFICATION

(This Page)

unclassified

(This Report)

unclassified

15. NUMBER OF PAGES

16. PRICE