PNNL-31245

Human Factors of Encoded Ultrasonic Examinations in Nuclear Power Plants

March 2021

T Sanquist
J Harrison

Prepared for the U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Under Contract DE-AC05-76RL01830
Interagency Agreement: NRC-HQ-60-17-D-00010

DISCLAIMER

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 Battelle Memorial Institute, 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, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

PNNL-31245

Human Factors of Encoded Ultrasonic Examinations in Nuclear Power Plants

March 2021

T Sanquist
J Harrison

Prepared for the U.S. Nuclear Regulatory Commission
under a Related Services Agreement with the U.S. Department of Energy
Contract DE-AC05-76RL01830
Carol A. Nove, Contracting Officer Representative

Pacific Northwest National Laboratory
Richland, Washington 99354

Abstract

This report presents an analysis of human factors that influence performance in automated/encoded ultrasonic examinations. Operating experience at two plants—Shearon Harris and Palisades—suggests that human factors contributed to missed detection of flaws that were clearly visible in subsequent review of recorded data. This report presents a task analysis of encoded ultrasonic examinations that illustrates the greater complexity of the work process as compared to manual examinations. Interviews with seven subject matter experts identified issues affecting human performance reliability, including knowledge and experience, the examiner's process, task complexity, equipment, team coordination and cohesion, and time pressure. These issues are illustrated with quotes and discussion. Review of the medical image interpretation literature—a surrogate task for encoded examination—revealed that human error occurs in 3.5–4% of all imaging studies, with perceptual misses and cognitive biases underlying the majority of errors. Cognitive biases, such as overreliance on prior reports, appear to have been involved in the Palisades event, while fatigue from long work hours and distraction from work conditions were associated with the Shearon Harris event. The research suggests a variety of work practice mitigations to reduce error potential, including double reading by independent reviewers, the provision of a dedicated, distraction-free space for analysis, additional perceptual training for examiners, and completion of interpretation without interruption.

Summary

The Nuclear Regulatory Commission (NRC) Office of Nuclear Regulatory Research (RES) is conducting research to identify the human factors associated with performance in nondestructive examination (NDE). The current research focuses upon automated/encoded ultrasonic testing (UT), as recent operating experience (OE) at the Shearon Harris and Palisades plants have shown that flaws in recorded data were not detected by examiners. NRC inspection reports and root cause analyses suggested that specific human factors such as long work hours, failure to compare current results with multiple prior year history reports, and lack of independent review contributed to the events.

The methods employed for this project are based on approaches developed and described by Sanquist et al. (2018), including task analysis by functional decomposition of the work process, interviews with subject matter experts (SMEs), identification of frequently recurring themes in the interview comments, and review of the scientific and technical literature that evaluates similar tasks.

Task analysis of encoded UT illustrates that these types of exams are more complex than manual UT, both in terms of the number of people involved and the equipment. Encoded UT is performed by teams consisting of acquisition and analysis personnel. In addition to the ultrasonic probes, automated UT utilizes a drive mechanism that is mounted on the component to be examined, an encoder to record position information, and displays for analysts to interpret data after it is acquired. Functional task modeling illustrates the additional tasks required to perform automated examinations, including equipment setup and subsequent data interpretation by the analyst team. Data are interpreted based on various views of the imagery, and comparative assessment can be accomplished by evaluating multiple views with amplitude time/distance gates and A-scan waveforms. The relatively more complex process of interpretation creates more potential areas for error to occur.

The human factors literature concerning NDE is sparse—there are no reports concerning automated/encoded data analysis, thus the dearth of existing NDE-specific research necessitates review of domains that are similar to NDE. Human performance in medical image interpretation provides a comparative basis for identifying factors that may affect examiner performance in NDE. This domain entails the psychological processes of comparison, interpretation, and decision making, and provides a robust empirical literature documenting a variety of human performance effects. The medical imaging literature supports analysis of error types and causes, the effects of experience upon performance, cognitive judgments in image interpretation, and the effects of fatigue and distraction, and can therefore provide ideas regarding mitigations that might apply in NDE.

Review of the medical image interpretation literature suggests that common performance influencing variables—experience, cognitive biases, fatigue, and distraction—exert predictable effects upon interpreting imagery and data that is similar to that encountered in encoded ultrasonic examinations. Errors do occur, and they are of specific types. Early studies (Garland 1959) showed that experienced radiologists will miss 30% of the positive indications on chest radiographs and will improperly interpret 2% of the negative images as showing pathological indications (false positive). This research found 30% disagreement on the same cases by different radiologists and 20% disagreement upon re-reading by the same physician.

More recent analyses show that human error occurs in medical image interpretation at the level of approximately 3.5–4% of all imaging studies (Berlin 2007). Analyses of errors suggest that approximately 42% of the errors are due to under-reading, i.e., perceptual misses, and the remaining 58% result from a variety of other cognitive biases (Kim and Mansfield 2014). More experience is associated with better performance, and fatigue and distraction can impair accuracy. Cognitive biases, such as satisfaction of search (early search termination), and satisfaction of report (overreliance on prior interpretation by others) can lead to misses or wrong interpretations. The findings reviewed from the medical image interpretation domain have direct applicability to encoded ultrasonic image interpretation, both in terms of characterizing errors and their causes, as well as suggesting practical mitigation approaches.

Interviews were conducted remotely with seven SMEs (four Level IIIs and three Level IIs) with a range of experience from 10 to 38 years. All were employed by vendors, with one of the Level IIIs previously employed by a utility. This sample covered three of the four major vendors conducting encoded examinations in U.S. nuclear power plants.

The interview transcripts were analyzed for thematic content by the process described in Sanquist et al. (2018). The following performance influencing factors (PIFs) represent the most frequently occurring themes within the data. Illustrative quotes are provided in the body of the report. The dominant themes are:

Knowledge and Experience: On-the-job know-how for the entire team of people responsible for the examination is emphasized as an important element of running a smooth exam, anticipating potential problems and taking steps to avoid them.

Examiner Process: The theme of examiner process focuses on understanding what the individual is seeing in the data and various means of ensuring confidence in the analysis (such as double reading).

Task Complexity: Encoded UT involves more equipment, more variables, and more people than manual UT, resulting in complex examination processes across multiple exam functions.

Equipment: Encoded UT involves more equipment of a more complex nature.

Team Coordination and Team Cohesion: Communication among team members and their ability to anticipate situations based on experience working together is perceived as a key element in how smoothly an exam proceeds.

Time Pressure: Time is a concern in conducting encoded examinations, but it may be perceived differently by crew members depending on their role and the relationship between the vendor and the utility.

The types of problems observed with encoded UT examinations at Palisades and Shearon Harris appear to be a result of specific human factors variables. The Palisades event entailed multiple years of relying on prior year reports of a supposed geometric indication that was actually a growing flaw, and the misconception that a flaw could not occur on the inner diameter of the nozzle. This event illustrates a variety of cognitive biases discussed in the literature review, including:

- Satisfaction of report – Abnormality is missed because of overreliance on a report from a previous examination.

- Framing – Tendency to interpret an abnormality in different ways depending on how a case is presented.
- Lack of knowledge – A finding is seen but attributed to the wrong cause because of a lack of knowledge on the part of the interpreter.

The practice of reviewing only recent history contributed to the inability to see a clear growth pattern that was evident when a longer time span of examinations was reviewed and compared.

The Shearon Harris event was attributed to long work hours and suboptimal inspection conditions. Fatigue and generally reduced human performance occurs with increasing time on-the-job, particularly during night shifts where circadian disruptions can exacerbate the effects of fatigue. Distraction may also have occurred due to suboptimal inspection and analysis conditions.

In general, encoded UT examinations are more complex and require higher levels of knowledge and greater team coordination and cohesion. These factors are linked, such that more knowledgeable team members are perceived as doing a better and more efficient job and are able to anticipate the requirements of complex examinations and the needs of other team members, thus contributing to better overall reliability. Focused attention during data evaluation and interpretation is enhanced by providing dedicated space for the work that is free of interruptions and other distractions.

Specific work practice approaches to enhancing encoded UT examination reliability are supported by the research and include:

- Double reading – double reading is used in medical image interpretation on a selective basis, and it is the standard for identifying discrepancies, particularly in complex cases. This approach is routinely implemented by one of the vendors interviewed as standard operating procedure. Encoded data are reviewed simultaneously by two analysts, and they are blind to each other's analysis until finished and comparison can be made.
- Dedicated analysis space – this practice was mentioned by most participants as being an essential need for the intensive cognitive work of data interpretation. The characteristics of such space include separation from other work crew members and a structure that reduces noise, suggesting a separate room with a door and controlled access such that only key personnel are admitted.
- Training and practice – experience is essential for building domain-specific knowledge. This can be gained by on-the-job training, use of practice samples, and potentially through perceptual learning with UT simulators. These approaches provide critical feedback to trainees.
- Work schedule/breaks – Long shifts and round-the-clock work is common in NDE. Research indicates that fatigue develops with longer time-on-task, and particularly during night work. Interventions for this problem entail taking breaks during long, monotonous tasks, and ensuring that schedules use forward rotation if crew are to be varied on shifts over a long outage (e.g., moving from day shift to afternoon to night).
- Complete a reading of a data set without interruptions – this practice was identified by one respondent as his individual standard procedure, in order to ensure that he devoted full attention to a set of data. If an interruption happened, he started over. The medical image interpretation literature suggests that physicians who are interrupted in the midst of reading

an image tend to spend less time on that image when they resume, and more discrepancies occur. Complete reading without distraction can reduce this potential.

This focused study of human factors in encoded UT examinations indicates the important role of various PIFs and work practices in achieving reliable data acquisition and interpretation. The proposed work practices can be considered as guidance for utilities and vendors, and they are supported by the research documented in this report.

Acronyms and Abbreviations

CT	Computed tomography
DOE	United States Department of Energy
EPRI	Electric Power Research Institute
MRI	Magnetic resonance imagery
NDE	nondestructive examination
NRC	Nuclear Regulatory Commission
OE	Operating experience
PIF	Performance influencing factors
PNNL	Pacific Northwest National Laboratory
SME	Subject matter expert
UT	Ultrasonic testing

Contents

Figures

Tables

1.0 Introduction

This report presents the application of the task analysis methods developed in prior work (Sanquist et al. 2018) to the assessment of human factors in encoded ultrasonic testing (UT). Encoded examination differs from manual UT in that the data are acquired by means of probes that are moved by automated drive mechanisms applied to the surface to be inspected. This process entails more equipment and larger teams of people than a manual exam.

Recent operating experience (OE) suggests the need for a better understanding of the human factors associated with encoded UT examination reliability. For example, in May of 2013, a review of UT data from the Shearon Harris Plant revealed a 0.26-inch indication that was missed during an inspection in 2012. A root cause analysis determined that the indication was challenging to detect and that there was little the licensee could have done directly to cause the analysts to miss the indication. However, a special inspection report (Nuclear Regulatory Commission 2013) noted that analyst working conditions, including tight quarters, noise, distraction and fatigue from long work hours may have contributed to the missed indication. Independent review of the data was not performed during this inspection, which might have identified the indication.

In November of 2018, Palisades found leakage during visual inspection of their reactor pressure vessel head. Further analysis of historical data revealed that the leak was caused by primary water stress corrosion cracking in a control rod drive mechanism penetration tube. The crack had grown over the course of 12 years and was missed by inspectors as they discounted the idea that inner diameter cracking was possible. Additional information provided by the Nuclear Regulatory Commission's (NRC) (2019) review of the vendor root cause analysis suggests that examiner training that focused on outer diameter flaws led to a biased "mind-set" among analysts, and that the practice of comparing current inspection data with the most recent past examination obscured substantial growth in the flaw. As with the Shearon Harris event, independent review was not conducted.

The root cause analyses of both events revealed that the flaw indications were represented in the recorded data, and that the procedures used were qualified to reveal such flaws. Human factors affected analyst performance, including the misconception that the indication for Palisades was caused by geometry on the inner diameter, and for Shearon Harris, suboptimal inspection conditions and fatigue resulting from long work hours. Other cognitive and work practice factors and potential mitigations are discussed in the NRC inspection reports (Nuclear Regulatory Commission 2013; Nuclear Regulatory Commission 2019). It is noteworthy that both events occurred during reactor head exams, which tend to receive more attention from the utility due to the potential impact on the outage schedule; this can lead to implicit or explicit time pressure among the examiners.

The purpose of the present analysis is to determine if there are other human factors influencing the performance and outcomes of encoded UT, to identify potential errors, and to provide a basis for identifying prospective work practice mitigations.

2.0 Structure of the Report

We describe the methods applied in the project, followed by results and discussion. A detailed review of relevant image interpretation literature is presented first, followed by results from task analysis. These include a narrative contrasting manual and encoded UT, functional decomposition of the tasks involved in performing encoded UT, and SME interview data. The final section discusses work practice changes that may enhance examination reliability. Appendix A contains the interview protocol used with SMEs.

3.0 Methods

This report builds upon methods that were developed and applied in previous work by the PNNL team and discussed in Sanquist et al. (2018). These methods include task analysis by functional decomposition of the work process, interviews with SMEs, identification of recurring themes in interview comments, and review of scientific and technical literature that evaluates similar tasks.

The SME interview protocol was modified to focus on the specific role of performance influencing factors (PIFs) during encoded exams, rather than jointly identifying tasks and PIF interactions as in the previous work (Sanquist et al. 2018); see Appendix A. This modification was made because the functional task model was already developed on the basis of the earlier work.

Interviews were conducted remotely and recorded using web conferencing software. The audio recording files were then transcribed using transcription software that produced a time-stamped text file of the entire interview. These text files were subsequently reviewed, ambiguities corrected, and the identity of the speaker inserted. This process was facilitated by the ability to replay the specific interview segment based on the time stamp within the transcribed file.

4.0 Literature Review

This section presents the literature review portion of the project, which focused on understanding issues from medical imaging studies that may be applicable to NDE. Prior reviews of literature concerning NDE and human factors have elucidated a variety of performance effects and shaping factors that influence reliability (D'Agostino et al. 2017).

The D'Agostino et al. (2017) review identified a number of factors influencing NDE performance within the context of an overall sociotechnical systems model comprising task characteristics, individual differences, group characteristics, physical environment, and organizational factors. Among the key findings from this review were the following:

- Variation in performance across a range of NDE examiners in diverse application domains such as nuclear and aviation
- The relationship of experience/expertise to performance reliability
- The impact of examiner cognitive processes such as early search termination and disregard of data
- Performance decrements due to fatigue
- The influence of attention on performance.

Reproduction of the search strategy of D'Agostino et al. (2017) revealed no additional pertinent articles with the exception of Agnisarman et al. (2019). These authors reviewed automation-assisted visual inspection technologies across domains such as bridge deck examination, pipeline, and underwater cable integrity, and suggested that a better understanding of examiner "sensemaking" processes of data interpretation is important for enhancing reliability, particularly as new visualization approaches are used.

While previous literature suggests key variables influencing NDE examiner performance, the empirical findings are sparse due to limited opportunities for research in the domains of interest. Since the reliability issues discussed in the introduction are based on examiner review of encoded data images, our strategy for the current analysis was to identify an application domain that involves analogous tasks of image-based data evaluation and interpretation and is supported by a robust empirical literature. It is presumed that lessons learned from domains with a more extensive empirical base will translate to work practices in NDE. The paucity of empirical research articles concerning NDE requires seeking related information from a domain with analogous task requirements, specifically, interpreting medical images (Drury et al. 1990; Harris 1969; Waite et al. 2019). This literature supports analysis of error types and causes, the effects of experience upon performance, cognitive judgments in image interpretation, and the effects of fatigue and distraction, and can therefore provide ideas regarding mitigations that might apply in NDE.

Review of the scientific and technical literature was accomplished through a web-based search using standard search engines such as Google Scholar, PubMed, Web of Science, PsycINFO and Scopus as well as broad-based open internet searches. Initially, searches for information concerning human factors in encoded UT yielded results that have already been discussed by D'Agostino et al. (2017). More focused searches using combinations of terms such as "human factors," "image interpretation," "human error," and "diagnostic imagery," indicated that the main body of literature that would be pertinent to the current research focuses upon human factors in clinical radiology. Key review articles for the 2010–2020 period were identified, and specific empirical studies were retrieved on the basis of the reference lists and findings of interest (this often resulted in foundational papers extending much further back in time). The material reviewed is a selective representation of a voluminous literature, and we focus on reports that (1) are most pertinent to elements likely to be involved in encoded UT as suggested by task analysis and interview data, (2) are representative of the basic issues associated with radiologic image interpretation, (3) provide background on fundamental psychological processes that underlie specific problem areas such as experience, cognitive bias, fatigue, and distraction, and (4) discuss potential mitigation strategies.

There are numerous similarities between nondestructive examination and medical diagnosis. Analyses of industrial inspection (Drury et al. 1990; Harris 1969) delineate processes of comparison (with a mental standard), interpretation, and decision making. Medical diagnosis also entails a complex decision process based on testing. Both processes involve the following functional elements applied to either patients or plant components:

- Review of history
- Physical examination
- Diagnostic testing with non-invasive technologies
- Interpretation of test results
- Communication of test results
- Development of a treatment or intervention plan, if necessary.

A full discussion of these process elements in medicine is provided by the National Research Council report entitled *Improving Diagnosis in Healthcare* (National Research Council 2015), which describes numerous examples of the complexities of diagnosis across a variety of medical specialties, including image interpretation. This report emphasizes the importance of understanding the entire system involved in medical diagnosis, since errors that occur, for example during testing or imaging procedures, can perpetuate to later interpretation, diagnosis, and treatment. Similar conclusions have been reached regarding errors in NDE (D’Agostino et al. 2017).

4.1 The Process of Medical Image Interpretation

There are many varieties of medical images, from straightforward 2-dimensional chest X-rays to more complex volumetric images provided by computed tomography (CT) and magnetic resonance imagery (MRI). The basic task of the radiologist is to view the image(s), and to detect and interpret abnormal indications. There have been numerous studies of how this process is executed, differences between novices and experts, and the details of volume image reviews (Waite et al. 2019). Theoretical accounts of medical image interpretation, based on observational and eye movement studies, generally involve elements of “holistic” and focal visual information processing that proceed to some extent in parallel with holistic impressions, guiding attention to areas that may warrant more detailed review (Sheridan and Reingold 2017). Research indicates that expert performance in medical image interpretation is “domain-specific,” meaning that the ability to correctly detect and diagnose disease from imaging studies does not translate to better performance in non-medical visual scanning tasks. Performance is also specific to sub-domains within radiology, such that expertise in chest radiograph interpretation does not directly transfer to interpreting mammograms, for example.

Volumetric image interpretation can involve “stack mode” processing in which motion is simulated by scrolling through sequential images searching for “pop out” lesions that stand out from the background, scanning each slice widely before moving on to the next, or “drilling,” in which the eyes are held in a constant location while scrolling through the depth plane—this is a variant of stack processing (Waite et al. 2019).

4.2 Error in Medical Image Interpretation

Studies of human error in medicine have documented the widespread occurrence of errors of many different types, including diagnosis. Current estimates suggest that 12 million adults experience a diagnostic error each year (Singh et al. 2014), including with medical imaging. Error in medical imaging has been systematically studied for over sixty years, starting with studies by Garland (1959). The early studies showed that experienced radiologists will miss 30% of the positive indications on chest radiographs and will improperly interpret 2% of the negative images as showing pathological indications (false positive). Garland’s research found 30% disagreement on the same cases by different radiologists and 20% disagreement upon re-reading by the same physician. Since the time of this research, many other studies have found essentially the same pattern across a diversity of imaging modalities (Berlin 2007). The miss and false positive rates translate to an overall error rate of 3.5 to 4% when the entire population of positive and negative images is used as a denominator.

The early research in studies of error in radiology defined error as based primarily on perceptual or other issues. Perceptual errors were initially estimated to occur at a rate of approximately 60–80% (Berlin 1996), with the remainder attributed to technique or cognitive factors such as lack of knowledge, biases, and other factors (Bruno et al. 2015). Retrospective studies of large numbers of errors (resulting from delayed diagnoses and difficult case conferences) have developed more detailed classification systems on the basis of multiple expert review (Kim and Mansfield 2014). Table 1 presents this classification system, definitions, and rate of occurrence of the errors (multiple errors could occur with a single case).

Table 1. Classification scheme for errors in diagnostic radiology (Kim and Mansfield 2014).

Cause of Error	Definition	Occurrence (%)
Overreading (false positive finding)	Finding is appreciated but attributed to wrong cause	0.9
Faulty reasoning	Finding is appreciated and interpreted as abnormal but is attributed to the wrong cause (true positive finding misclassified)	9.0
Lack of knowledge	Finding is seen but attributed to the wrong cause because of a lack of knowledge on the part of the interpreter	3.0
Under-reading (missed finding)	Finding is present on the image but is missed	42.0
Poor communication	Abnormality is identified and interpreted correctly, but the message does not reach the clinician	0.0
Technique	Finding is missed because of the limitations of the examination or technique	2.0

Cause of Error	Definition	Occurrence (%)
Prior examination	Finding is missed because of failure to consult prior radiologic studies or reports	5.0
History	Finding is missed because of inaccurate or incomplete clinical history	2.0
Location	Finding is missed because of the location of a lesion outside the area of interest on an image	7.0
Satisfaction of search	Finding is missed because of failure to continue to search for additional abnormalities after the first abnormality was found	22.0
Complication	Complication from a procedure	0.5
Satisfaction of report	Finding was missed because of overreliance on the radiology report from a previous examination	6.0

The largest proportion of errors in this study were perceptual (42%), i.e., a missed clinical finding because it was simply not seen; the study's authors attribute this to under-reading and suggest that checklists may be appropriate to reinforce active search patterns. Errors resulting from failure to continue the search after finding an abnormality—"satisfaction of search"—accounted for 22%. This is a common cognitive bias (discussed below), sometimes described as "when a finding is made, the thinking stops." Faulty reasoning accounted for 9% of the errors, and the study authors noted the contribution of lack of experience in the interpreting radiologist. A related error type—lack of knowledge—accounted for 3%; together, these experience-based errors accounted for 12% of the findings. An unusual location of the lesion was associated with 7% of the errors, in areas such as the periphery of the image, and often related to scrolling through a series of volume images and missing an abnormality on the first or last view. Errors that were perpetuated across imaging studies—termed "satisfaction of report"—accounted for 6% and are based on the influence of a prior radiologist's interpretation on the current review; this occurs when the radiologist reads a previous report and is unduly influenced by its contents. Other categories of error occur due to failure to consult previous studies, inaccurate clinical history, faulty technique for image acquisition, and false positives.

4.2.1 Relevance to NDE

The error patterns described above are also apparent in NDE. Perceptual errors, i.e., misses or false positives are reported in studies of inspector performance, with a substantial proportion of results falling outside the optimal performance range, as well as problems with sizing for smaller defects (Harris and McCloskey 1990). More detailed analysis of error types was reported by Harris and McCloskey (1990) from a study of defect detection using verbal protocol transcripts for analysis of cognitive processes during examination. The results showed that detection errors (either misses or false positives) were associated much more frequently with reaching early conclusions and disregarding evidence thought to be irrelevant. These types of decision errors are similar to "satisfaction of search" in radiological image error analysis.

4.3 Factors Influencing Error in Medical Image Interpretation

Analysis of error and other human factors studies of image interpretation have identified a number of factors that influence performance, including experience, specific cognitive biases, radiologist fatigue, and workload-based distractions. Each of these factors is discussed below.

4.3.1 Experience Effects

Development of expertise to interpret radiological images is like many skilled professions—a result of experience and practice. Studies of medical expertise have shown that simple accrual of time is not sufficient for developing expert skill; rather, it is a function of frequency of practice on challenging cases representative of the domain (Ericsson 2018). While the error analysis discussed above noted a role of experience and knowledge in the cases studied, there have been relatively few head-to-head comparisons of novice and expert performance outside of eye movement studies (discussed below).

Direct comparison of image interpretation performance for attending physician mammographers, residents, and technologists was reported by Nodine et al. (1999). In this study, pairs of images—a proportion of which contained malignant lesions—were presented to the physicians, and decision time and accuracy were recorded. The overall results of receiver operating characteristic analysis showed that attending physicians were significantly better than residents and technologists, who were not statistically different. There was a significant relationship between the number of cases that had been previously read by each group, such that more case experience (attending physicians) was associated with higher levels of accuracy. Decision speed was faster for the attending physicians (15.6 sec) than for residents (21.6 sec) and technologists (28.1 sec). The authors interpret the results as follows:

> "…experts are perceptually more sensitive in recognizing lesions than are those with less expertise because the experts have read more mammogram cases, seen more lesions, and differentiated more lesions into malignant and benign categories. In practical terms, this means that through massive amounts of experience experts become perceptually tuned to recognizing familiar breast structures and detecting odd or novel variations in them." (p. 584)

In practice, attending physician mammographers read approximately 10,000 cases over a three-year period, whereas residents in a mammography rotation may read about 650 cases. Nodine and Mello-Thoms (2019) suggest that radiology expertise is subdomain-specific, such as learning how to read chest X-rays, which entails a substantial component of peripheral vision, does not transfer to improving skills for interpreting mammograms.

A large-scale study of mammography performance and physician experience was reported by Miglioretti et al. (2009) involving over 1.5 million images and 250 physicians. The general finding was similar to that of Nodine et al. (1999), i.e., more experience resulted in fewer patient recalls for resolution of ambiguity, and false positive findings decreased. The results suggested that specific fellowship training in mammography was associated with better performance earlier in the career span.

Considerable effort has been devoted to studies of eye movements in radiological image interpretation, since the technique can show the scanning processes involved and the patterns of skill acquisition with experience (Waite et al. 2019). One of the earliest studies of this type was reported by Kundel and La Follette (1972), in which the eye movement patterns of staff radiologists were compared with medical students and residents. The findings suggested that the staff radiologists had developed characteristic scanning that was not seen in medical students and that had begun to develop in residents by the fourth year of training (Figure 1).

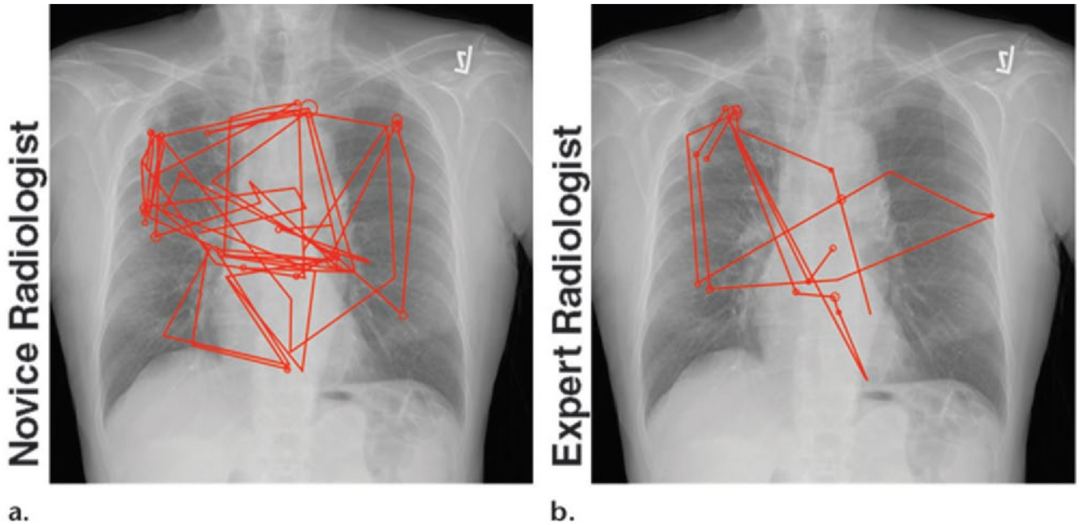


Figure 1. Eye movement scanning patterns for novice (a) and expert (b) radiologists. (Drew et al. 2013).

It is generally found that over time, experts develop an efficient pattern of eye movements for particular types of images and are often able to tell at a single glance that an abnormality is present (Drew et al. 2013). While there is considerable variability across viewing patterns for experts, they do show faster movement to the site of abnormalities and fewer movements to areas less likely to be diagnostic. This type of research also shows that areas falsely rejected as negative, or falsely judged as positive are reflected by increased visual dwell times that are as long or longer than gaze duration at abnormality-containing areas; true positives, false positives, and false negatives all show longer visual dwell times than true negative decisions (Krupinski 1996; Kundel et al. 1989; Nodine et al. 2002).

The training of radiologists emphasizes comprehensive and orderly search patterns for specific types of images. However, most empirical studies using eye movement tracking have failed to show superior performance of students trained in systematic viewing, although they may cover a larger portion of the image while viewing (Aufferman et al. 2015; Kok et al. 2016; van Geel et al. 2017); see (Waite et al. 2019) for review). Experts are thus thought to use a variety of “free search” patterns that are guided by holistic impressions, focal attention, and systematic coverage guided by experience (Waite et al. 2019).

4.3.2 Experience Mitigations

Mitigation for errors resulting from lack of experience include case conferences and feedback to trainees as part of a traditional residency program. Fellowships—available at selected institutions and for high-performing specialization candidates—can provide further exposure to the high volumes of image interpretation necessary to develop expertise.

Simulators for ultrasound and interventional radiology procedures are beginning to be employed in medical training programs. Tolsgaard et al. (2015), for example, reported a study of physician residents who were trained conventionally (i.e., by supervised ultrasound performance on live patients) compared to a group trained clinically whose learning was examined using an ultrasound simulator. Follow-up assessment of both groups indicated that the physicians trained with the ultrasound simulator performed better on all measures of an objective examination assessment and maintained their skills for several months beyond the training period.

A variety of other systems are available for simulation in the procedural skills associated with neuroradiology and cardiac catheterization, for example, but simulation has not had a substantial role in training image interpretation skills. This process is generally still accomplished through a master-apprentice model. There have been localized implementations of interpretation skill simulation modules, but they are not standardized or widely available (Cook et al. 2016) due to lack of faculty time, expertise for implementation, and lack of perceived need.

A more limited approach to simulation-based experience can potentially be accomplished with lower-fidelity perceptual learning modules. Perceptual learning refers to “experience-induced improvements in the pickup of information” (Kellman and Massey 2013). Numerous laboratory tasks have shown that with experience, subjects are able to more quickly discover information that is relevant to their task domain and also become more fluent, i.e., they can extract that information with greater speed and ease. This might be considered a form of learning to “see” with greater proficiency—as expertise develops, the ability to discern patterns in complex visual stimuli improves over novices who cannot discern such patterns. A variety of application-oriented studies suggests that relatively simple instruction on image interpretation, followed by classification of large numbers of images, can lead to novices performing nearly as well as experts for radiographs, histopathology slides, and electrocardiograms (Kellman 2013).

4.3.3 Relevance to NDE

Experience has been identified as an important variable in ultrasonic examiner performance (D’Agostino et al. 2017). Relationships similar to those shown in radiology, i.e., increasing performance accuracy with increasing expertise, have not been demonstrated to our knowledge. Instead, the importance of experience is emphasized in various interview and survey studies, with the relationship presumed to be similar to that shown in other areas of expert performance. Studies of performance demonstration qualification and re-qualification show that experienced examiners can fail at rates higher than less experienced examiners with recent refresher training (Stephens 2000). The counter-intuitive result of less experienced examiners exhibiting better detection accuracy is most likely a result of an intensive three-week training taken prior to testing. This lack of expertise-accuracy relationship is also likely an effect of small sample sizes in the studies that have been conducted, as well as the emphasis on quality for the junior examiners. It is probable that with larger study populations and consistent treatments across groups a small but positive relationship between expertise and accuracy would be demonstrated.

The implications of the expertise-accuracy relationship for NDE are clear—greater amounts of training and experience in interpreting encoded ultrasonic imagery will yield better examiner performance. The means to achieve this are the same as in radiology—specialized training, extensive practice, and (potentially) the use of low-fidelity simulation. The Electric Power Research Institute (EPRI) is developing a low-fidelity ultrasonic training simulator, and when coupled with the ability to incorporate data from field studies and to insert artificial flaws, it may help to compensate for the relative lack of field-based training experience that has been described in other studies (Sanquist et al. 2018).

4.4 Cognitive Judgment Effects

Interpreting radiographic images involves making judgments under uncertain conditions—essentially a series of decisions that entail such factors as prior probability of disease conditions, clinical history provided with the imaging test order, sub-specialty of the physicians and the conditions they most frequently see, and typicality or atypicality of the abnormality presentation. To facilitate the many decisions that humans make every day, two modes of thought have evolved (Kahneman 2011), referred to generically as System 1 (fast, intuitive) and System 2 (slow, deliberate). All humans use both modes, but System 1 dominates, particularly in circumstances that are familiar. These modes of thinking are involved in diagnostic decisions and System 1 thinking entails a variety of "heuristics" or rules-of-thumb to facilitate rapid information processing and pattern recognition. Heuristics include the tendency to judge things that are similar as belonging to the same category, placing more weight on examples of things that easily come to mind, and the tendency to seek information that confirms initial impressions. When heuristics lead to erroneous decisions, they are known as cognitive biases. Table 2 lists a number of cognitive biases that occur in rapid decision making and how they can affect a diagnostic outcome. Reviews of diagnostic errors across many different medical specialties (National Research Council 2015) indicate that cognitive bias is a widespread issue.

Table 2. Cognitive biases and heuristics that can affect examiner reliability. Adapted from Ely et al. (2011).

Cognitive Bias/Heuristic	Description
Anchoring	Tendency to perceptually "lock-on" to salient features of the image too early in the diagnostic process and failure to adjust this impression in light of later information.
Availability	Tendency to judge things as being more likely or frequently occurring if they readily come to mind.
Confirmation bias	Tendency to seek information that supports initial impression.
Framing	Tendency to interpret an abnormality in different ways depending on how case is presented.
Base rate neglect	Tendency to ignore the true prevalence of an abnormality, either inflating or reducing its base rate and distorting likelihood.
Satisfaction of search/ Premature closure	Decision making process ends too soon; the diagnosis is accepted before it has been fully verified. "When the diagnosis is made, the thinking stops."
Satisfaction of report	Abnormality is missed because of overreliance on a report from a previous examination.
Representativeness constraint	Interpreter looks for prototypical manifestation of abnormalities (pattern recognition) and fails to consider atypical variants.
Unpacking principle	Failure to elicit all relevant information in establishing a differential diagnosis.
Context errors	Critical signal is distorted by the background against which it is perceived.

Review of Table 2 suggests that the heuristics and biases are related; satisfaction of report and satisfaction of search, for example, both result in abbreviated searches and reasoning processes. The interaction between fast and slow modes of thought depends on a variety of factors, including the familiarity of the circumstances, and the required pace and workload. There are various theories of how these systems interact (National Research Council 2015); one influential theorist (Kahneman 2011) suggests that System 1 fast thinking generates a "default" response, which may be overridden by System 2 slow and deliberate thinking based on the situation (for example, conflicting opinions in a case conference review).

The detailed study of errors by Kim and Mansfield (2014) suggests that satisfaction of search and satisfaction of report account for 26% of the cognitive errors classified. Examples of cognitive bias in radiology include:

- Satisfaction of search/premature closure: Missing a secondary hand fracture after review of a frontal plane radiograph showed a finger fracture. The hand fracture was visible only on an oblique radiograph; search terminated with the initial fracture (Itri and Patel 2018).
- Availability bias: A patient presented with a 10-day history of abdominal pain and mild fever. Ultrasonic images were reviewed by a radiologist who had recently given a lecture on tumor formation; an abdominal tumor was diagnosed. As the fever and white blood count increased, further imaging revealed features that led to the correct diagnosis of ruptured appendicitis with a pelvic abscess (Busby et al. 2018).
- Satisfaction of report: A female patient with a history of breast cancer and lumpectomy was re-examined per follow-up protocol. Prior reports described non-malignant scarring, which was confirmed by mammogram. Subsequent evaluation with additional imagery and pathologic analysis led to the correct diagnosis of recurrent malignancy (Busby et al. 2018).

Busby et al. (2018) present further examples of cognitive bias errors with associated images, and initial and revised diagnoses.

4.4.1 Cognitive Judgment Mitigations

A variety of approaches to mitigate perceptual error cognitive bias have been proposed, including:

- Checklists
- Cognitive forcing/debiasing or "reflective" thinking
- Double reading.

Checklists have been introduced to reduce error in high-risk proceduralized work such as surgery. Evaluation studies suggest that compliance is variable and that reduction of adverse events requires consistent usage (Pugel et al. 2015). In radiology, the process of image interpretation is initially based on fast, intuitive System 1 thinking, and checklists for this type of activity are problematic with respect to potentially altering the expert eye movement patterns developed over time (Ganeshan et al. 2018). Checklist-type structured reports have been introduced for various types of image interpretation, with the largest success being in breast imaging. Studies of structured reports for other imaging types (e.g., abdominal, cervical spine) have shown that clinically significant findings are increased by a substantial percentage (see (Ganeshan et al. 2018), for review). Large-scale surveys of structured report usage suggests that they are used in some form in approximately 50% of institutions internationally. A variety of issues related to (1) potentially altering the perceptual process, (2) clumsy user interface, and

(3) poor templates, combine to reduce acceptability within the radiology profession (Ganeshan et al. 2018). Improvements in technology over time are anticipated, which may address some of these technology acceptance issues.

Approaches to reflective thinking, based on studies of "metacognition" (thinking about thinking) have been proposed to slow down or counter the heuristics and biases resulting from the fast and intuitive process of interpretation. These include teaching modules that train students to recognize when they might be subject to cognitive bias, and self-reflective questions to ask during or after image interpretation (e.g., "Have I considered alternative diagnoses?"). While students are able to identify potential biases after training, there is no evidence to support interventions based on cognitive debiasing and reflective thinking (see (Norman et al. 2017) for review of multiple studies).

Double reading is generally considered the gold standard for identifying discrepancies and reducing errors in radiological image interpretation (Garland 1959). Studies evaluating various forms of double reading (higher level specialist, same level specialist, or sub-domain specialist) have found that accuracy is generally increased across the board, with the best-value trade-off in time being for sub-domain specialist review for selected high-risk cases (Geijer and Geijer 2018). In practice, peer review is implemented in the United States for 5% of the cases in each department, per requirements of the Joint Commission on Accreditation of Healthcare Organizations (Chetlen et al. 2020). Implementation of double reading for all imagery is infeasible due to workload and cost issues (Waite et al. 2017), but if done on a selective basis can improve diagnostic and treatment outcome.

4.4.2 Relevance to NDE

Cognitive bias effects have been observed in detection testing conducted with ultrasonic examiners (Harris and McCloskey 1990). In a study involving 26 examiners conducting a total of 235 pipe weld inspections, low detection accuracy was associated with several biases listed in Table 2, including satisfaction of search, satisfaction of report, anchoring, and confirmation bias. In a follow-up study (Harris 1992), a simple checklist decision aid was provided to 43 examiners who conducted 257 weld inspections. The performance of examiners with the decision aid was 25% better than examinations conducted by a group using a standard approach. The key elements of the decision aid involved delaying a final decision until all data had been gathered and prompts for specific types of signal characteristics to evaluate.

Double reading has been reported in interview data (see Interview Results, Section 6.0 of this document), but it is not commonly employed by all vendors. One vendor reported double reading of all encoded data as a standard practice; other individual examiners reported consulting a colleague if in their opinion a particularly complex or ambiguous set of results warranted a second review.

4.5 Fatigue Effects

There exists an extensive scientific literature exploring the relationship between fatigue and adverse outcomes in work settings. In occupational studies, aspects of work schedules are examined as preconditions for fatigue-related adverse outcomes such as error, injury, and death. In some cases, work performance is measured in the field, either as an indicator for fatigue or as a proxy for presumed risk of an accident or incident. Measurement of fatigue itself is problematic; it is not directly observable and must be inferred from other, measurable phenomena. Fatigue is the presumed mechanism linking long periods of wakefulness and performance-related adverse events (Williamson et al. 2011). Therefore, the effect of work schedule on safety and effective job performance is hypothesized to operate primarily through prior sleep–wake history and the time of day at which work takes place.

Cognitive indicators of fatigue can include degraded alertness and attention, problems with sustained concentration, tendency to be easily distracted, confusion, forgetfulness, memory problems, and performance worries. Psychomotor and cognitive speed, vigilant and executive attention, working memory, and higher cognitive abilities appear to be particularly affected by sleep loss (Lim and Dinges 2008). These cognitive decrements can accumulate to severe levels over periods of chronic sleep restriction without the full awareness of the affected individual (Van Dongen et al. 2003). Effect indicators can include demotivation (such as boredom, lack of desire and enthusiasm, or temporary feelings of depression) and coping, emotional, or interactional fatigue (such as anxiety, avoidance, comfort seeking, irritability, or feeling stressed) (Kamdar et al. 2004; Luna et al. 1997). These effects show considerable variability across individuals (Van Dongen et al. 2005). Microsleeps, sleep attacks, and lapses in cognition are considered to be an indication of state instability (i.e., short duration transitions between sleep and wake states) (Doran et al. 2001). The following effects of fatigue are generally agreed upon in the scientific literature (Caldwell et al. 2008):

- Accuracy and timing degrade;
- Attentional resources are difficult to divide;
- A tendency toward repetitive behavior patterns develops;
- Social interactions decline;
- The ability to logically reason is impaired;
- Attention wanes;
- Attitude and mood deteriorate; and
- Involuntary lapses into sleep begin to occur.

In medicine, most studies focus on the effects of very long, overnight shifts of 24 hours or more, which are typical of physician training programs; such schedules have been demonstrated to result in sleep loss among medical interns (Barger et al. 2005; Lockley et al. 2004). Studies of medical staff in the field and in the lab have shown an association between sleep loss and impaired job performance (Weinger and Ancoli-Israel 2002).

Similar results have been observed in radiology. After a day of reading, accuracy degrades and subjective fatigue increases for both conventional radiography and CT interpretations (Krupinski et al. 2012). A study of satisfaction of search (Krupinski et al. 2017) suggests that with increasing fatigue, a more conservative criterion for reporting is adopted, such that true positives and false positives are reduced. Hanna et al. (2018) report a study of image interpretation following a night shift showing reduced detection accuracy, which was associated with longer visual scan times and longer times to fixate on the abnormality when it was present.

4.5.1 Fatigue Mitigations

There are relatively few countermeasures available for fatigue other than work schedule interventions to reduce long hours and circadian rhythm disruptions. Waite et al. (2017) suggest combating fatigue with naps, appropriate light exposure, optimized ergonomics such as workstation lighting, reduced physical stressors, social interaction, and caffeine. These are the standard countermeasures often suggested by researchers; caffeine and planned naps have been shown to be relatively practical to implement and have been shown to temporarily reduce the performance losses from fatigue (Sanquist et al. 2018; Smith-Coggins et al. 2006). Despite the recognition of fatigue as a pervasive problem within healthcare systems, it remains a difficult issue to address across the many different occupations involved, including radiology (Kancherla et al. 2020).

4.5.2 Relevance to NDE

Fatigue can reduce the accuracy of ultrasonic inspection as a result of the time-on-task, sleep deprivation, circadian rhythm effects and a combination of all three. Performance of NDE inspection has been studied primarily by Drury and colleagues using simulations of aircraft component inspection tasks (summarized in Drury and Watson (2002)). These researchers have not found performance decrements that might be expected on the basis of a voluminous psychological literature demonstrating reduced accuracy over time (Poulton 1973). Drury and Watson suggest that despite negative evidence of a vigilance decrement in their studies, it is prudent to believe that such an effect occurs and to structure jobs in such a way as to reduce the potential impact:

> "It would be safest to assume that some vigilance decrement potentially exists for all inspection tasks, but that many conditions can prevent this from affecting actual performance. Such conditions can include good feedback, social interaction, high signal rate or accurate briefing on defect expectations….Where conditions are unfavorable (rare signals, long time on task, low feedback, etc)….it would be prudent to limit the period of continuous inspection."(p. 31)

Folkard and Tucker (2003) conducted an analysis of industrial efficiency studies and safety incidents, and combined data to calculate relative risks. They concluded that safety risks (injury or performance decrement) increase from morning to afternoon to night shifts, with night shifts being 30% more risky than morning shifts. Similarly, over the course of a night shift, risk appears to be greatest in the second hour (11:00 p.m.–12:00 a.m.), decreasing over subsequent hours of the work period. Risk also increases on each successive night in schedules where workers are assigned to multiple night shifts. For example, risk is 36% higher on the fourth successive night of a multiple night shift schedule.

The preponderance of evidence—from studies of medical image interpretation, vigilance and time-on-task experiments, and risk analyses of industrial efficiency and safety—suggests that similar impacts will occur in NDE and the interpretation of encoded ultrasonic data.

4.6 Distraction Effects

Interpreting medical images is a task that demands focused attention. There are many variables that affect the degree to which attention is concentrated (Kahneman 1973; Pashler et al. 2001). Distraction is often considered the opposite of attention—some external or internal information that reduces the ability to focus on elements of the environment that are important for job performance. Examples of external distractions include ringing mobile phones, ad-hoc interruptions by staff, loud conversations, and sudden changes in the scene being observed. Internal distractions can involve thinking about something that is not related to the task at hand, such as family or money issues, or can involve getting very highly focused on some aspect of a task (trying to find patient clinical history) to the exclusion of other elements of image interpretation (concentrated focus on abnormality identification). It has been proposed that "inattention" is a result of internal distractions and that external distractions prevent the effective maintenance of attention (Regan et al. 2011). In most situations, these distinctions are not entirely clear-cut. For example, simply carrying a mobile phone that is powered on while interpreting images can constitute either an internal distraction ("I wonder when he will call") and an external distraction, such as a text message arriving while reviewing image details. From a practical standpoint, distraction can be considered as a demand on attention that reduces focus on the task of primary importance for effective job performance and safety.

Mobile technologies such as smart phones and iPads are increasingly common in medical settings. Various advantages have been enumerated, including rapid order entry to central medication systems, local/personal access to medical record information, imagery viewing, and time savings over workstation-based approaches (Halamaka 2011). A survey study of smartphone usage during inpatient rounds at a New York academic hospital (Katz-Sidlow et al. 2012) found that the majority of clinicians (residents, house staff, and faculty) use personal smartphones in their work. This survey also reported that more than 20% of the house staff self-reported missing significant clinical information during rounds due to smartphone distraction. Further, observations of other staff suggest that the proportion may be even higher (as great as 40%). For example, interruptions from a text message received by a resident as he was entering a critical medication change order resulted in the order not being completed and subsequent adverse health consequences to the patient (Halamaka 2011). The increasing trend in organizations is for a "bring your own device" policy, which reduces cost but provides connectivity benefits. However, interruptions from personal messages or calls in critical circumstances can have extremely adverse effects.

Interruptions and distractions including phone calls and emails can lead to diagnostic error in medical image interpretation. Balint et al. (2014) evaluated the association of discrepant interpretations with phone calls received during a shift and found that accuracy of interpretation was affected by phone calls received in the hour preceding the generation of a discrepant report. An experimental study using a secondary task (Wynn et al. 2018) showed that distraction led to increased reading time and decreased accuracy for subtle cases for both residents and attending physicians. An on-call radiologist may be interrupted an average of two and a half times at peak hours while interpreting a CT study (Yu et al. 2014). Eye tracking research suggests that interruption by phone calls can reduce the amount of time spent viewing images, compared to similar cases that are not interrupted (Drew et al. 2018).

4.6.1 Distraction Mitigations

Many authors advocate reducing interruptions and distractions if possible (Bruno et al. 2015; Busby et al. 2018; Waite et al. 2017). However, some interruptions may be helpful, so rather than eliminating interruptions altogether, Waite et al. (2017) suggest using clinical assistants and caller identification systems to help prioritize and pace distractions and interruptions, especially during active image interpretation and other error-prone clinical scenarios. Levy et al. (2020) report an evaluation of employing a call triage assistant—a trainee staff member, to initially field and screen calls to the radiology reading room. Their analysis showed that using such an assistant to handle routine, non-medical inquiries reduced interruptions by 71% and increased image interpretation turn-around time by 30 minutes; interpretation accuracy was not affected.

4.6.2 Relevance to NDE

Maintaining focused attention on sensitive tasks is a requirement of many jobs and everyday activities, and it is particularly important in tasks involving visual search for subtle signals that are infrequent—one of the basic tasks of NDE. We have not found any specific studies of distraction effects upon the reliability of NDE. However, the parallels between the many studies of driver distraction and impacts upon medical image interpretation suggest that any circumstances that reduce examiner focused attention while reviewing data can potentially reduce performance accuracy. The most well-studied situation that is analogous to NDE data interpretation is driving; drivers continuously scan the visual environment for signs, signals, and dangerous situations. Distractions occur in the form of phone calls, in-vehicle information systems, and passengers—all of which have been shown to impair driving performance (Strayer et al. 2011). It is thus likely that similar sources of distraction during NDE task performance will degrade accuracy.

Prior interview studies (Sanquist et al. 2018) suggest that distraction occurs during manual ultrasonic examination, based on health physics personnel communicating with examiners during critical tasks. In the current study, considerable emphasis was placed upon having a dedicated data analysis area that is free from distraction, rather than set up in a common area where people are coming and going. One respondent indicated that if he was interrupted during his review of encoded data, he started again from the beginning.

The empirical data from medical imaging studies, driver distraction research, and the interview reports from the current and prior interview studies performed in the NRC's NDE research program all suggest that providing a distraction-free environment will enhance the quality of data interpretation.

4.7 Summary of Literature Review Findings

Review of the medical image literature suggests that common performance influencing variables—experience, cognitive biases, fatigue and distraction—exert predictable effects upon interpreting imagery and data that is similar to that encountered in encoded ultrasonic examinations. Errors do occur, and they are of specific types. More experience is associated with better performance, and fatigue and distraction can impair accuracy. Cognitive biases, such as satisfaction of search, can lead to misses or wrong interpretations. The findings reviewed from the medical image interpretation domain have direct applicability to encoded ultrasonic image interpretation, both in terms of characterizing errors and their causes, as well as suggesting practical mitigation approaches.

5.0 Task Analysis Results

A primary method for studying human factors in encoded UT examination is task analysis. This approach was described and applied to manual UT by Sanquist et al. (2018). The main techniques used for the encoded UT analysis were functional decomposition of the work process and SME interviews with content analysis.

5.1 Description and Contrast of Manual and Encoded UT Examinations

This section provides a general description of the functions and tasks used for manual and encoded UT, contrasting specific aspects of data acquisition and analysis. The basis for this description is the prior manual UT task analysis (Sanquist et al. 2018), an overview of UT evaluation and imaging by Crawford et al. (2015), and the functional decomposition of encoded UT based on EPRI procedure (EPRI-ENC-DMW-PA-1)—described in the subsequent section. The basic functional model of encoded UT is portrayed in Figure 2, which shows that a fundamental difference from manual UT is the inclusion of a Data Interpretation and Evaluation function.

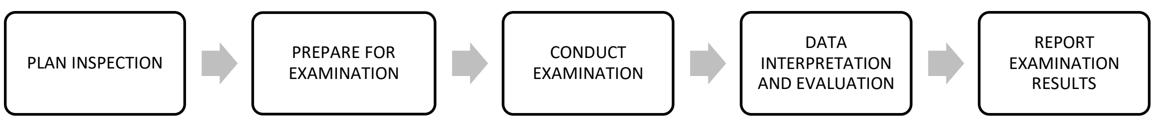


Figure 2. Functional structure of encoded UT examination.

5.1.1 Manual Conventional UT Examination

In the manual conventional UT process, the examiner simultaneously collects and interprets signals, and adapts the procedure according to what is observed. Initially, the examiner may conduct a full sweep with the UT probe of the area of interest and “watch the roll” of the resulting waveform on the scope. Large deflections that are not attributable to weld or pipe geometry are noted for more detailed assessment. Once the entire surface to be scanned is completed thoroughly, the examiner can return to potential indications for further data collection and evaluation. This process is driven by the procedure-specific flowchart and may involve additional probe angles, linear measurements on the pipe/weld surface, photography of the area, notes regarding location and measurement, and if available, stored images of the waveforms on the scope. Once outside the examination area, additional data can be used, as necessary, to determine the significance of the observed indications, including past examination reports, radiographs, and other NDE results as available.

The acquisition process and the need to conduct evaluation of indications during the data collection process is a primary difference from the encoded UT process. If appropriate evaluation data are not collected, due to health physics concerns, probe availability, or other factors, an ambiguous situation would occur, potentially confounding interpretation and/or requiring additional data collection in the area of interest.

5.1.2 Encoded UT Examination

Encoded examinations differ fundamentally from manual UT in how data are acquired. An instrumentation setup process involves technicians applying the encoder, probe, and drive mechanism to the surface to be inspected. A key element of this process is ensuring that data quality and coverage will be acceptable. This determination is generally a consensus decision made among the instrumentation technicians who set up the equipment, and the more senior analysts. Once the instrumentation setup is deemed acceptable, the scan is initiated and quality monitored during acquisition.

Data interpretation and evaluation in the encoded UT process is conducted “off-line,” i.e., it is done after the data are collected (with the exception of a quality check described above). Substantially more data are available to the analyst, including the ability to display the raw UT waveform as would be seen on a manual conventional display, as well as various “views” of the examination volume. These include (1) a view of the weld as if looking down from the top, similar to viewing a radiograph of a weld, (2) a view of the weld from top to bottom, stretched out over the entire surface (i.e., the pipe is essentially “unfolded” or “unrolled” so that a linear view of the entire interior circumference is provided), and (3) a side view, providing a visualization of the long-axis (an axial view). Additionally, the raw A-scan waveform is provided in a separate section of the display. These various views are set up by analysts individually, but there is some overall consistency in how they are employed. The use of specific amplitude and time/distance gates can be used to declutter or enhance displays, as required, to distinguish flaws from geometric reflectors. Cursors can be placed at specific areas of interest in order to measure depth and size of flaws and to correlate observations among the various views of the signal, as well as with other supplemental data (exam history, radiographs, etc.). The display views are linked via software, so if the analyst scrolls along the circumferential axis, the corresponding sections of the other views are updated and displayed.

The potential for error in encoded analysis comes from different factors than manual convention UT. This includes the improper setting of amplitude gates on one or more of the interpretive displays, issues of maintaining proper orientation, i.e., top and bottom of pipe on a stretched circumferential view, distinguishing inner and outer diameter, and properly transcribing location data from the display during reporting.

5.2 Functional Decomposition

A more detailed analysis of the functions, tasks, and subtasks for encoded UT is presented in Table 3. This structure is the same as that used for manual UT functional decomposition. Certain tasks are altered for encoded UT—these are shown in red italicized font. Manual UT task elements, such as moving and skewing the probe by hand have been removed.

Table 3. Functional decomposition of encoded UT examination. Differences from manual UT are shown in red italicized font.

Function	Task		Subtask	
1. Plan for Examination	1.1	Develop plan based on requirements and previous data	1.1.a	Review and revise procedures
	1.2	Identify qualified examiners	1.2.a	Bid contract if necessary

	1.3 Ensure component prepared for examination in the field	1.3.a Walk-down areas to be examined 1.3.b Determine special equipment, personnel, and access requirements 1.3.c Measure components and access routes 1.3.d Prepare work packages
2. Prepare for Examination	2.1 Receive plant orientation and training	
	2.2 Review work package and procedure	2.2.a Obtain work package 2.2.b Review drawings, OE, location of component, weld history, weld profile, previous scanning technique, etc. *2.2.c Review prior exam results* *2.2.d Check system linearity and channel function (element check)* 2.2.e Develop scan plan according to procedure *Raster or line scan (circumferentially oriented flaws)* *Raster scan (axially oriented flaws)* *Select manual or fully automated drive* *Select reference system associated with technique* *Create data acquisition layout to allow operator to observe and monitor data acquisition* *Calculate focal laws for probes* *Conduct channel function check* *Create examination setup file detailing parameters of inspection* 2.2.f. Determine recording criteria for indications according to procedure
	2.3 Verify that the component is within the procedure and personnel qualification ranges	
	2.4 Assemble materials and equipment. *Test equipment functionality prior to staging*	2.4.a Select scope, transducers, cables, couplant gel, *drive mechanisms*, etc. according to procedure 2.4.b Re-familiarize self with equipment *2.4.c Verify data software dataset*
	2.5 Calibrate equipment to procedure	2.5.a Save calibration files 2.5.b Complete calibration forms
	2.6 Attend pre-job briefing	2.6.a Ask questions during pre-job briefing, as needed
	2.7 Coordinate examiner responsibilities	
	2.8 Pack equipment and materials	
	2.9 Prepare to enter area for exam	2.9.a Walk to area of plant where component is located/entrance to radiological-controlled area 2.9.b Obtain radiation work permit

		2.9.c Remember allowable dose and radiation levels in work area 2.9.d Meet with Health Physics to review radiation work permit 2.9.e Obtain dosimetry devices 2.9.f Dress out for radiation and industrial environment as appropriate 2.9.g Enter radiological-controlled area
	2.10 Locate component	2.10.a Monitor dose at regular intervals 2.10.b Monitor surroundings for hazards
3. Conduct Examination	3.1 Verify at correct component and weld	
	3.2 Verify conditions at weld	3.2.a Verify accessibility of component for scanning 3.2.b Document physical surroundings that would prevent 100% examination scan coverage (e.g., nearby structures, insulation, component geometry, etc.) 3.2.c Verify surface preparation 3.2.d Document surface conditions 3.2.e Ensure conditions at weld meet procedural requirements 3.2.f Decide if examination can proceed
	3.3 Perform supplemental inspection measurements	3.3.a Measure weld crown width, temperature, contour, etc. as required by procedure
	3.4 Set up for scan	3.4.a Identify scan area (may include drawing marks at weld center line or additional locations) 3.4.b Verify area to be scanned (examination volume coverage) 3.4.c Assess flow direction 3.4.d Determine scan start position (may include the use of markings) 3.4.e Verify scan plan (ordering of scans; with flow/against flow; clockwise/counterclockwise)
	3.5 Perform calibration verification ("cal check")	3.5.a Select calibration file for first search unit 3.5.b Verify calibration with reference block
	3.6 *Mount scanner drive mechanism*	*3.6.a Ensure correct probe placement is perpendicular to weld for circumferential flaws (axial scan) or parallel to weld for axial flaws (circumferential scan)* *3.6.b Perform probe position calibration* *3.6.c Perform pre-exam verifications* *3.6.d Start couplant system* *3.6.e Adjust instrument scan sensitivity*
	3.7 Write down notes (and take pictures in the field)	

	3.8 Verify data is complete	*3.8.a Data analyst to ensure coverage of Code required examination volume* *3.8.b Data analyst to ensure quality is acceptable*
	3.9 Wipe off gel and clean area (pack up in field)	3.9.a Pack up equipment and materials 3.9.b Clear equipment with health physics 3.9.c Remove protective clothing 3.9.d Return to work area
4. Data Interpretation and Evaluation of Indications	4.1 Verify post-examination calibration	
	4.2 Flaw Detection Analysis	*4.2.a Gate volume to identify benchmark information* *4.2.b Gate regions within volume. Compare patterns and responses from different wave modes, angles, skews and beam directions* *4.2.c Adjust sensitivity (gain or palette) to improve contrast as necessary*
	4.3 Classify indications as flaw or geometry *4.4 Determine size and depth per procedure* *4.5 Record non-relevant indications in sufficient detail to assist future examinations.*	
5. Report Examination Results	5.1 Complete examination documentation	5.1.a Complete calibration data sheet (form) 5.1.b Analyze results 5.1.c Complete examination data sheets (forms) 5.1.d Document indications and limitations as needed
	5.2 Perform post-job briefings	5.2.a Communicate potential indication of flaw if appropriate 5.2.b Review extenuating circumstances with job coordinator

5.2.1 Functional Decomposition – Discussion

The task model delineated in Table 3 involves more functions, tasks, and subtasks than manual UT. In particular, the development of a more detailed scan plan is necessary (subtask 2.2.e) to ensure proper coverage and allow examiners to monitor data acquisition. Additionally, the scanner mechanism needs to be mounted and the probe placements verified and calibrated (Task 3.6, subtasks 3.6 a-e). Prior to initiating the full scan, a short segment of pipe is scanned to verify setup and data quality are acceptable (subtasks 3.8 a-b). The most significant difference between manual and encoded UT is the addition of the Data Interpretation and Evaluation of Indications function (Function number 4). The tasks and subtasks within this function are conducted "off-line" after all data are collected as described above.

The functional task model shown in Table 3 is more complex than manual UT, and also involves more people. The encoded UT team consists of a lead examiner who does the overall preparation of the plan and performs the data interpretation function, as well as technician teams of two or more individuals who are responsible for setting up and running the equipment while conducting the exam to acquire data. Encoded UT examinations are thus more complex from a functional task standpoint, and also more involved in terms of personnel.

6.0 Interview Results

SME interviews were conducted with seven UT examiners whose participation was solicited by email from PNNL. The seven examiners included four Level IIIs and three Level IIs, with experience ranging between 10 and 38 years. All were employed by vendors, with one of the Level IIIs having also been previously employed by a utility. The interviewees represented three of the four major vendors that currently conduct encoded examinations.

Table 4 shows the summary results from the PIF coding process, and the number of interview transcripts in which that PIF was mentioned. The frequency of comments should be interpreted as a relative gauge of the salience of that particular factor in the overall conduct of encoded examinations. The number of interview transcripts in which a PIF was mentioned shows the consistency of responses across the seven interviewees. In the following paragraphs, we discuss those PIFs that received ten or more comments across interviews, as that appears to be a meaningful breakpoint between higher and lower frequency, both for numbers of comments and how often a PIF was mentioned across interviewees. Review of the comments that were mentioned less than ten times does not reveal any substantial differences with how they were discussed in the manual UT task analysis (Sanquist et al. 2018).

Table 4. Number of comments for each PIF and number of interviews where PIF is mentioned at least once.

Performance Influencing Factor	Number of Comments	Number of Interview Transcripts
Knowledge/Experience	29	7
Examiner Process	28	6
Task Complexity	25	6
Equipment	17	5
Team Cohesion	15	5
Time Pressure	11	5
Team Coordination	11	6
Pre-Job Preparation	8	4
Workload/Stress/Fatigue	7	3
Accessibility of Component	7	4
Utility Planning	6	2
Industry Challenges	6	2
Vendor-Utility Interaction	4	3
Procedure	3	1
Training	3	3
Cognitive Factors	1	1
Lighting	1	1

Performance Influencing Factor	Number of Comments	Number of Interview Transcripts
Radiation	1	1
Supervision	1	1
Motivation/Attitude/Personality	0	0
Physical Ability	0	0
Noise	0	0
Temperature/Humidity	0	0
Organizational Culture	0	0

6.1 Content Analysis Themes

The themes identified in the data were organized by PIF to characterize examiners' perspectives on the specific dimensions within these factors that influence performance. Each factor listed in this section includes a general integrating statement reflecting the overall theme. This is followed by a discussion of the dimensions of that factor as reflected in the data and presentation of selected examples of comments from the interviews that illustrate themes associated with that factor.

Knowledge and Experience: On-the-job know-how for the entire team of people responsible for the examination is emphasized as an important element of running a smooth exam, anticipating potential problems, and taking steps to avoid them.

The most frequently mentioned PIF was knowledge and experience—not only for the lead examiner/data analyst, but also for the technicians responsible for setting up the equipment on the components to be examined. This factor is linked in the interview comments to issues of higher task complexity for encoded examinations, and to team coordination/cohesion. The basic theme reflects the importance of experience gained through performing multiple exams in realistic settings and learning the details of equipment set up that are appropriate to the circumstances at hand (sometimes the equipment setup is referred to as "tooling" or "track setting"). Selected comments regarding Knowledge and Experience include:

- *the real learning curve with an encoded exam is … learning the parameters of the system and how that system relates to the exam. In other words, if you have a transducer on a particular side of the weld and you're looking at the weld, what is that probe rotational angle when you go in, it's setup. So that's really the difficult part about an encoded exam. It's learning the system and how to input data into that system correctly and how to run the system.*
- *some examiners just have a better ability to keep in mind all of those things that may actually contribute to, say, potential artifacts in the data or other issues that may give you something that isn't precisely what you think you're seeing.*

Examiner Process: The theme of examiner process focused on understanding what the individual is seeing in the data and various means of ensuring confidence in the analysis.

The modified interview protocol addressed the function of Data Interpretation and Evaluation of Indications, and the comments reflected this. Most analysts described how they interact with the different display views, techniques of "gating" that can be used to distinguish flaw indications from geometry and to cause isolated indicators to "pop out," relating current observations to prior examinations, and getting additional opinions from other analysts. Comments also reflected the lead examiner's role in reviewing equipment setup and data quality, but that is covered in discussion of Knowledge and Experience above. Selected comments for this theme include:

- *I really want to try to look at as much of the scan as I can, if I can look at the entire full range of the scan, if that was the 360 degrees around the pipe or a component. I want to try to have that on my image first just to look at everything and make sure that base metal noise levels and things are pretty much the standard and the same around the entire component.*
- *You may want to review radiographs. If you see something odd, something you want to verify...You may want to run a scan again, occasionally trying a supplemental probe or to just try and get better data. You might want to, for instance, do a high resolution scan at slower speed to try and get more and better information.*
- *It gets tough if it's in the middle of interpretation with interruptions. We usually run a process of how we analyze where if you stop in the middle my rule of thumb is to start over again. If I have to stop to get a status report, I don't skip steps, I start over.*
- *By gating in, closing your gates and searching through, it helps to separate something to pop out that's different.*
- *Most times there's a minimum of three people that look at this exam data before we accept it or reject it.*

Task Complexity: Encoded UT involves more equipment, more variables, and more people than manual UT, resulting in complex examination processes across multiple exam functions.

The complexity of encoded UT is based on multiple factors—equipment, exam acquisition variables, and the people involved in setting up, acquiring, and interpreting the data. Encoded UT involves more equipment (multiple rolling duffels full), essential variables, and technicians (tooling/track setters and lead data analysts). As a consequence, there is considerable attention to set up details and verification to ensure that data are acquired properly and interpretation is based on a correct understanding of the set up variables (e.g., index and scan axes are correct, data phase is correct). Additionally, communication between team members inside and outside of containment is necessary to evaluate data quality prior to full component scan data acquisition. Selected comments regarding task complexity include:

- *If you don't have a good crew mounting your equipment to keep it all straight, you're just not going to get a good exam out of it.*
- *Another way that encoded varies from manual UT is there are different people with different responsibilities providing information for that exam. For example, when the encoder device is mounted at the component, positioning data of the probe has to be given to the operator so that he can input that correctly. You don't have that with a manual exam. So, if you don't get that information correct, it's not fed from team member A, team member B to be input correctly, your result is going to be an inadequate exam.*
- *There are a number of errors that I recall having seen. For example, setting the scanning device up wrong and scanning the weld backwards—that can occur and has actually*

happened more than once. If that happens, all of the data are 180 degrees out of phase. I've also seen where the probe rotation angles that I mentioned before were not set properly. They were incorrectly input. So that throws the data off. With some encoded systems, that's readily apparent in the way that it will display the data, but on others, it's not.

Equipment: Encoded UT involves more equipment of a more complex nature.

The greater amount of equipment for encoded exams involves additional staging prior to setting up for an exam. This involves unpacking equipment that has been shipped, checking the functions, and also setting up the many essential variables and data views—most of which can be taken from previously stored values. Tight physical spaces can create problems for attaching an encoder device, potentially affecting exam coverage.

- *Aside from just having adequate space available on the component to perform the exam, you need to ensure you have adequate space for the encoding device. Depending on the exam, the physical size of that device could vary.*
- *One of the biggest headaches with encoded equipment are breakdowns and failures with the encoded device because they get a lot of work and they're usually treated pretty rough, by moving them around a lot. Things break on them. It will slow the schedule.*
- *Our phased array exams are usually done with software that has the capability of saving layouts. You can save whatever layout you prefer to make it as personal as you want. And you can do that ahead of time, because usually we have previous data that we review before the job…typically when I get the data, I can just open it up, select my layout and it loads everything up. My gates are already set for me and it's all taken care of pretty much as long as you set it up ahead of time.*

Team Coordination and Team Cohesion: Communication among team members and their ability to anticipate situations based on experience working together is perceived as a key element in how smoothly an exam proceeds.

Team Coordination and Cohesion are discussed together as the comments illustrate the linkage of communicating between the larger crews involved in encoded UT, and the advantages of teams that are familiar with each other. Both PIFs were mentioned with identical frequency in the coded comments. The comments reflect both the need for more overt communication in encoded UT due to the complexity of the tasks, and the desirability of developing shared team awareness of work practices that can facilitate the exam, since some elements of procedure can be anticipated based on past experience.

- *When you're doing an automated, encoded examination, you have to rely a lot on a whole team of people, not necessarily just yourself. We rely on people setting up the tooling. We rely on people acquiring the data. We all have to be on the same page, and we brief and brief and brief it.*
- *We find the people that work best together and continue to put them on the same projects. It makes everything go so much better. For example, your acquisition person—you have worked with them before and it is someone you know is good. It really helps. They do things before they need telling… they can anticipate potential issues that you might otherwise have to ask them to take care of...the team works together almost unconsciously since they've done it so many times before together.*
- *With automated exams, if you're not doing it regularly with a set team, it just makes things more difficult.*

Time Pressure: **Time is a concern in conducting encoded examinations but may be perceived differently by crew members depending on their role and the relationship between the vendor and the utility.**

Time pressure is a common issue in plant outages, and it is perceived in various ways by the interviewees. The comments generally did not reflect time pressure as a universal and intolerable problem, but instead one that can develop under certain circumstances such as equipment breakage or work within a highly contaminated area. Two interviewees stated that they had never felt specifically pressured; others described situations in which plant personnel were obviously in a hurry to get things done.

- *If a delay is encountered, then that could domino the schedule, which could put a lot of time pressure on the examiners.*
- *Whether data quality is acceptable is another big issue that there's debate on...quite often… Particularly when you're under pressure to get things done.*
- *They'll give us our separate area now (for analysis). And there's as long as you need (for interpretation).*
- *We've had no problem with customers pressuring anybody on the analysis of the data.*
- *With automated exams, typically, they're more specialized exams. A lot of times, they come into critical path. There's a lot more time pressure. Usually the customer will have somebody assigned to keep a close eye on the progress of the exam. We have task leads, for the job that are supposed to be the in-between and kind of be the block for us, that they deal with the customer. With these high-pressure times, on critical path jobs, the task lead deals directly with the customer.*
- *I've actually been in situations where you had plant managers standing outside the door asking you every five minutes, are you done? We have to start this plant up. That's a lot of pressure when you're the one guy holding them starting that plant up. And I'm serious. I have seen some unruly pressure. So, you rush it up and maybe you see something, and you wonder is that a flaw? Is that geometry? Well, nothing was called into previous data. I'm being rushed here. I don't think it's anything rather than actually taking the time to maybe say, hey, we need to go back in and look at this one specific area to take someone with some wisdom.*

Five comments across three different interviewees mentioned the importance of a separate dedicated space that is free from distractions for data analysis. Although this was not a specific PIF, and is related to potential distractions (which is a work stressor), we believe it is important to describe because it can have a bearing on data analysis reliability and suggests a straightforward work practice that can be employed when it is planned for. Specific comments regarding dedicated workspace for analysis include:

- [*For data analysis] you generally get a pretty good spot where you're not distracted. Sometimes in the past, you'd get put in a trailer with a dozen other people or more, making it tough to concentrate.*
- *Typically, in certain exams, it is written in as a requirement to prevent distractions [a separate area for analysis]. Usually what we'll do is put up a stop sign on the door. If we're put in a trailer, we'll put a stop sign on the trailer door with a phone number. You are not to enter.*

7.0 Error Modeling

As with the task analysis of manual conventional UT examinations, the functional decomposition and interview results provide a basis for analyzing potential error pathways and their consequences. The data suggest that there are no material differences between manual and encoded exams in potential errors in the functions of Planning, Conducting, or Reporting functions. We did obtain results suggesting several potential types of error for the Preparing for Examination and the Data Interpretation and Evaluation functions, as shown in Tables 5 and 6.

Table 5. Potential precipitating factors, error types, and consequences for the Preparing for Examination function in encoded UT.

Selective Precipitating Factors	Error Type	Potential Consequences
	Preparing for Examination	
Analyst does not confirm acceptable data and probe contact in pre-scan	Incomplete data due to loss of couplant or probe liftoff in small areas	Flaw may not appear in data and will not be detected
Reversing index and scan axes in setup	Data image out of phase	Locations of indications incorrect and misinterpreted
Skewed encoder mounting	Response characteristics may be present outside of the expected region; incomplete coverage	Missed flaw detection

Table 6. Potential precipitating factors, error types and consequences for the Data Interpretation and Evaluation function in encoded UT.

Selective Precipitating Factors	Error Type	Potential Consequences
	Data Interpretation and Evaluation	
Analyst does not confirm 100% acceptable data and probe contact	Incomplete data due to loss of couplant or probe liftoff in small areas	Flaw may not appear in data and will not be detected
Analyst does not follow all procedural steps to evaluate UT image response	Systematically dismissing an indication as geometry without a complete evaluation	Flaw mischaracterized as geometry or another anomaly
Overreliance on previous data analysis	Analyst accepts previous interpretation of data	Flaw mischaracterized as geometry or another anomaly
Interruptions or distraction during data interpretation	Less attention devoted to data upon resumption of task	Missed flaw detection
Inspection gate selection too tight	Response characteristics may present associated responses outside of the expected region	Missed flaw detection

The error types and precipitating factors illustrated in Tables 5 and 6 are not meant to be a comprehensive listing; instead, we present them as was done for manual UT—as illustrations of an error analysis framework that could be extended through additional analysis and to facilitate development of mitigations. While the types of errors listed in these tables are plausible and have been observed by the interview respondents, they also tend to be prevented by standard work practices such as use of stored files and procedural check points.

8.0 Conclusions & Recommendations

The research documented in this report—literature review, task analysis, and subject matter expert interviews—suggests that the types of problems observed with encoded UT examinations at Palisades and Shearon Harris are a result of specific human factors variables. The Palisades event entailed multiple years of relying on just the prior year report of a supposed geometric indication that was actually a growing flaw, and the misconception that a flaw could not occur on the inner diameter. This event illustrates a variety of cognitive biases discussed in the literature review, including:

- Satisfaction of report – Abnormality is missed because of overreliance on a report from a previous examination
- Framing – Tendency to interpret an abnormality in different ways depending on how a case is presented
- Lack of knowledge – a finding is seen but attributed to the wrong cause because of a lack of knowledge on the part of the interpreter.

It also appears that the practice of reviewing only recent history contributed to the inability to see a clear growth pattern that was evident when a longer time span of examinations was reviewed and compared.

The Shearon Harris event was attributed to long work hours and suboptimal inspection conditions. Fatigue and generally reduced human performance occurs with increased time on-the-job, particularly during night shifts where circadian disruptions can exacerbate the effects of fatigue. Distraction in the form of suboptimal inspection conditions can further erode focused attention that might otherwise result in detection of flaws.

At a more general level, the research reported here indicates that encoded UT examinations are more complex and require higher levels of knowledge and greater team coordination and cohesion. These factors are linked, such that more knowledgeable team members are perceived as doing a better and more efficient job and are able to anticipate the requirements of complex examinations and the needs of other team members, thus contributing to better overall reliability. The need for focused attention during data evaluation and interpretation is enhanced by providing dedicated space for this work that is free of interruptions and other distractions.

Review of the medical image interpretation literature suggests that common performance influencing variables—experience, cognitive biases, fatigue and distraction—exert predictable effects upon interpreting imagery and data that is similar to that encountered in encoded ultrasonic examinations. Errors do occur, and they are of specific types. Early studies (Garland 1959) showed that experienced radiologists will miss 30% of the positive indications on chest radiographs and will improperly interpret 2% of the negative images as showing pathological indications (false positive). This research found 30% disagreement on the same cases by different radiologists, and 20% disagreement upon re-reading by the same physician.

More recent analyses show that human error occurs in medical image interpretation at the level of approximately 3.5–4% of all imaging studies (Berlin 2007). Analyses of errors suggest that approximately 42% of the errors are due to under-reading, i.e., perceptual misses, and the remaining 58% result from a variety of other cognitive biases (Kim and Mansfield 2014). More experience is associated with better performance, and fatigue and distraction can impair accuracy. Cognitive biases, such as satisfaction of search, can lead to misses or wrong interpretations. It is noteworthy that the interview results did not reveal a substantial impact of cognitive factors or bias. This is likely because biases are generally unconscious processes, and independent review is necessary to identify them. The findings reviewed from the medical image interpretation domain have direct applicability to encoded ultrasonic image interpretation, both in terms of characterizing errors and their causes, as well as suggesting practical mitigation approaches.

The consolidated results of this research suggest several approaches to work practice in encoded UT examinations that can enhance the reliability of the process. These practices include:

- Double reading – this approach is routinely implemented by one of the vendors interviewed as standard operating procedure. Encoded data are reviewed simultaneously by two analysts, and they are blind to each other's analysis until finished and comparison can be made. Double reading is used in medical image interpretation on a selective basis and is the standard for identifying discrepancies, particularly in complex cases.
- Dedicated analysis space – this practice was mentioned by most participants as being an essential need for the intensive cognitive work of data interpretation. The characteristics of such space include separation from other work crew members and a structure that reduces noise, suggesting a separate room with a door and controlled access such that only key personnel are admitted.
- Decision aiding and checklists – reduction of cognitive bias errors in image interpretation has not been successful in medicine, but this may be due to the use of very weak instruments that are not specific to image interpretation. The one study of decision aiding available for NDE (Harris 1992) suggests that a simple checklist that reminds examiners of key aspects of the A-scan—such as rise time, the signal "walk," and persistence with probe skew—can facilitate flaw detection with subtle indications. This type of more focused analysis is often built into procedure flowcharts for characterizing flaws.
- Training and practice – the research reported here is commensurate with the voluminous literature concerning expertise, i.e., experience is essential for building domain-specific knowledge. This can be gained by on-the-job training, use of practice samples, and potentially through perceptual learning with UT simulators. These approaches provide critical feedback to trainees. The medical imaging literature does not specifically address feedback, but the structure of residency and fellowships are designed to provide it through oversight and review.
- Work schedule/breaks – long shifts and round-the-clock work is common in NDE. Research indicates that fatigue develops with longer time-on-task, and particularly during night work. Interventions for this problem entail taking breaks during long, monotonous tasks, and ensuring that schedules use forward rotation if crew are to be varied on shifts over a long outage (e.g., moving from day to afternoon to night). Alternatively, selection of personnel who experience fewer difficulties during night work would be appropriate if such individuals can be identified.
- Complete a reading of a data set without interruptions – this practice was identified by one respondent as his individual standard procedure in order to ensure that he devoted full attention to a set of data. If an interruption happened, he started over. Generally, respondents reported being able to review data from an entire weld within 90 minutes. The medical image interpretation literature suggests that physicians who are interrupted in the midst of reading an image tend to spend less time on that image when they resume, and more discrepancies occur. Complete reading without distraction can reduce this potential.

The empirical and literature review research reported here indicates that the increased complexity of encoded UT exams, coupled with specific PIFs, can lead to error. The types of errors observed in NRC event reports have parallels within the medical imaging literature and appear to be based on the same cognitive and situational variables. A variety of mitigations are available, as discussed above. An area that warrants additional research is the development of a decision aid specific to encoded UT that may help to reduce cognitive biases. Development of UT simulators for training and practice is also an area worthy of further work, as the research shows the clear influence of knowledge and experience on reliability in this complex task.

9.0 References

Agnisarman, S, S Lopes, KC Madathil, K Piratla, and A Gramopahye. 2019. "A survey of automation-enabled human-in-the-loop systems for infrastructure visual inspection." *Automation in Construction* 97: 52-76.

Aufferman, WF, BP Little, and ST Tridadapani. 2015. "Teaching search patterns to medical trainees in an educational laboratory to improve perception of pulmonary nodules." *Journal of Medical Imaging* 3 (1): 011006 - 011006-011005.

Balint, BJ, SD Steenburg, L Hongbu, C Shen, JL Steele, and RB Gunderman. 2014. "Do telephone call interruptions have an impact on radiology resident diagnostic accuracy?" *Academic Radiology* 21: 1623–1628.

Barger, L. K., B. E. Cade, N. T. Ayas, J. W. Cronin, B. Rosner, F. E. Speizer, C. A. Czeisler, Harvard Work Hours, Health, and Safety Group. 2005. "Extended work shifts and the risk of motor vehicle crashes among interns." *N Engl J Med* 352 (2): 125-134. https://doi.org/10.1056/NEJMoa041401.

Berlin, L. 1996. "Perceptual errors." *American Journal of Roentgenology* 167: 587-590.

Berlin, L. 2007. "Accuracy of diagnostic procedures: Has it improved over the past five decades?" *American Journal of Roentgenology* 188: 1173–1178.

Bruno, MA, EA Walker, and HH Abujudeh. 2015. "Understanding and confronting our mistakes: The epidemiology of error in radiology and strategies for error reduction." *Radiograhics* 35: 1668–1676.

Busby, L. P., J. L. Courtier, and C. M. Glastonbury. 2018. "Bias in radiology: The how and why of misses and misinterpretations." *Radiographics : A Review Publication of the Radiological Society of North America, Inc* 38 (1): 236–247.

Caldwell, J. A., J. L. Caldwell, and R. M. Schmidt. 2008. "Alertness management strategies for operational contexts." *Sleep Med Rev* 12 (4): 257-273. https://doi.org/10.1016/j.smrv.2008.01.002.

Chetlen, A. L., J. Petscavage-Thomas, R. A. Cherian, A. Ulano, S. B. Nandwana, N. E. Curci, R. T. Swanson, R. Artrip, T. K. Bathala, L. M. Gettle, and L. A. Frigini. 2020. "Collaborative Learning in Radiology: From Peer Review to Peer Learning and Peer Coaching." *Acad Radiol* 27 (9): 1261-1267. https://doi.org/10.1016/j.acra.2019.09.021.

Cook, T. S., J. Hernandez, M. Scanlon, C. Langlotz, and C. D. Li. 2016. "Why Isn't There More High-fidelity Simulation Training in Diagnostic Radiology? Results of a Survey of Academic Radiologists." *Acad Radiol* 23 (7): 870-876. https://doi.org/10.1016/j.acra.2016.03.008.

Crawford, SL, MT Anderson, AA Diaz, MR Larche, MS Prowant, and AD Cinson. 2015. "Ultrasonic evaluation and imaging." In *Integrated Imaging and Vision Techniques for Industrial Inspection*, edited by Z Liu, H Ukida, P Ramuhali and K Niel, 393–412.

D'Agostino, A, S Morrow, C Franklin, and N Hughes. 2017. *Review of human factors in nondestructive examination.* Nuclear Regulatory Commission. ML17059D745.

Doran, S. M. , H. P. A. Van Dongen, and D. F. Dinges. 2001. "Sustained attention performance during sleep deprivation: Evidence of state instability." *Archives Italiennes de Biologie* 139 (3): 253–267.

Drew, T., K. Evans, M. L. Vo, F. L. Jacobson, and J. M. Wolfe. 2013. "Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images?" *Radiographics* 33 (1): 263-274. https://doi.org/10.1148/rg.331125023.

Drew, T., L. H. Williams, B. Aldred, M. E. Heilbrun, and S. Minoshima. 2018. "Quantifying the costs of interruption during diagnostic radiology interpretation using mobile eye-tracking glasses." *J Med Imaging (Bellingham)* 5 (3): 031406. https://doi.org/10.1117/1.JMI.5.3.031406.

Drury, C, and J Watson. 2002. *Good practices in visual inspection.* FAA Technical Report. https://www.faa.gov/about/initiatives/maintenance_hf/library/documents/media/human_factors_maintenance/good_practices_in_visual_inspection_-_drury.doc.

Drury, CG, P. Prabhu, and A Gramopadhye. 1990. "Task analysis of aircraft inspection activities: Methods and findings." Proceedings of the Human Factors Society 34th Annual Meeting.

Ely, JW, ML Graber, and P Croskerry. 2011. "Checklists to reduce diagnostic error." *Academic Medicine* 86 (3): 307 – 313.

Ericsson, KA. 2018. "The differential influence of experience, practice, and deliberate practice on the development of superior individual performance of experts." In *The Cambridge Handbook of Expertise and Expert Performance*, edited by A Kozbelt AM Williams, KA Ericsson, and RR Hoffman, 745–769. Cambridge: Cambridge University Press.

Folkard, S, and P Tucker. 2003. "Shift work, safety and productivity." *Occupational Medicine* 53: 95–101.

Ganeshan, D., P. T. Duong, L. Probyn, L. Lenchik, T. A. McArthur, M. Retrouvey, E. H. Ghobadi, S. L. Desouches, D. Pastel, and I. R. Francis. 2018. "Structured Reporting in Radiology." *Acad Radiol* 25 (1): 66-73. https://doi.org/10.1016/j.acra.2017.08.005.

Garland, LH. 1959. "Studies on the accuracy of diagnostic procedures." *American Journal of Roentgenology* 82 (25-38).

Geijer, H., and M. Geijer. 2018. "Added value of double reading in diagnostic radiology,a systematic review." *Insights Imaging* 9 (3): 287-301. https://doi.org/10.1007/s13244-018-0599-0.

Halamaka, J. 2011. "Order Interrupted by Text: Multitasking Mishap." PS Net. Agency for Healthcare Research and Quality. https://psnet.ahrq.gov/webmm/case/257.

Hanna, T. N., M. E. Zygmont, R. Peterson, D. Theriot, H. Shekhani, J. O. Johnson, and E. A. Krupinski. 2018. "The Effects of Fatigue From Overnight Shifts on Radiology Search Patterns and Diagnostic Performance." *J Am Coll Radiol* 15 (12): 1709-1716. https://doi.org/10.1016/j.jacr.2017.12.019.

Harris, DH. 1992. *Effect of decision making on ultrasonic examination performance.* TR-100412. Palo Alto: Electric Power Research Institute.

Harris, DH, and BP McCloskey. 1990. *Cognitive correlates of ultrasonic inspection performance.* NP-6675. Palo Alto: Electric Power Research Institute.

Harris, Douglas H. 1969. "The Nature of Industrial Inspection." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 11 (2): 139-148. https://doi.org/10.1177/001872086901100207.

Itri, J. N., and S. H. Patel. 2018. "Heuristics and cognitive error in medical imaging." *American Journal of Roentgenology* 210 (5): 1097–1105.

Kahneman, D. 1973. *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Kahneman, D. 2011. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kamdar, B. B., K. A. Kaplan, E. J. Kezirian, and W. C. Dement. 2004. "The impact of extended sleep on daytime alertness, vigilance, and mood." *Sleep Medicine* 5: 441–448.

Kancherla, BS, R Upender, JF Collen, and et al. 2020. "Sleep, fatigue and burnout among physicians: an American Academy of Sleep Medicine position statement." *Journal of Clinical Sleep Medicine* 16 (5): 803–805.

Katz-Sidlow, R.J., A. Ludwig, S. Miller, and R. Sidlow. 2012. "Smartphone use during inpatient attending rounds: Prevalence, patterns, and potential for distraction." *Journal of Hospital Medicine* 7 (8): 595–599.

Kellman, P. J. 2013. "Adaptive and perceptual learning technologies in medical education and training." *Mil Med* 178 (10 Suppl): 98-106. https://doi.org/10.7205/MILMED-D-13-00218.

Kellman, PJ, and CM Massey. 2013. "Chapter Four - Perceptual Learning, Cognition, and Expertise." In *Psychology of Learning and Motivation*, edited by BH Ross. Cambridge, MA: Academic Press.

Kim, Y. W., and L. T. Mansfield. 2014. "Fool me twice: delayed diagnoses in radiology with emphasis on perpetuated errors." *AJR Am J Roentgenol* 202 (3): 465-470. https://doi.org/10.2214/AJR.13.11493.

Kok, E. M., H. Jarodzka, A. B. de Bruin, H. A. BinAmir, S. G. Robben, and J. J. van Merrienboer. 2016. "Systematic viewing in radiology: seeing more, missing less?" *Adv Health Sci Educ Theory Pract* 21 (1): 189-205. https://doi.org/10.1007/s10459-015-9624-y.

Krupinski, E. A., K. S. Berbaum, R. T. Caldwell, K. M. Schartz, M. T. Madsen, and D. J. Kramer. 2012. "Do long radiology workdays affect nodule detection in dynamic CT interpretation?" *J Am Coll Radiol* 9 (3): 191-198. https://doi.org/10.1016/j.jacr.2011.11.013.

Krupinski, E. A., K. S. Berbaum, K. M. Schartz, R. T. Caldwell, and M. T. Madsen. 2017. "The Impact of Fatigue on Satisfaction of Search in Chest Radiography." *Acad Radiol* 24 (9): 1058-1063. https://doi.org/10.1016/j.acra.2017.03.021.

Krupinski, Elizabeth A. 1996. "Visual scanning patterns of radiologists searching mammograms." *Academic Radiology* 3 (2): 137-144. https://doi.org/10.1016/s1076-6332(05)80381-2.

Kundel, H. L., and P. S. La Follette, Jr. 1972. "Visual search patterns and experience with radiological images." *Radiology* 103 (3): 523-528. https://doi.org/10.1148/103.3.523.

Kundel, HL, CF Nodine, and EA Krupinski. 1989. "Searching for lung nodules: visual dwell indicates false-positive and false-negative decisions." *Investigative Radiology* 24: 472 – 478.

Levy, J. L., C. W. Freeman, J. K. Cho, O. Iyalomhe, and M. H. Scanlon. 2020. "Evaluating the Impact of a Call Triage Assistant on Resident Efficiency, Errors, and Stress." *J Am Coll Radiol* 17 (3): 414-420. https://doi.org/10.1016/j.jacr.2019.11.007.

Lim, J., and D. F. Dinges. 2008. "Sleep deprivation and vigilant attention." *Ann N Y Acad Sci* 1129: 305-322. https://doi.org/10.1196/annals.1417.002.

Lockley, S. W., J. W. Cronin, E. E. Evans, B. E. Cade, C. J. Lee, C. P. Landrigan, J. M. Rothschild, J. T. Katz, C. M. Lilly, P. H. Stone, D. Aeschbach, C. A. Czeisler, Health Harvard Work Hours, and Group Safety. 2004. "Effect of reducing interns' weekly work hours on sleep and attentional failures." *N Engl J Med* 351 (18): 1829-1837. https://doi.org/10.1056/NEJMoa041404.

Luna, T., J. French, and J. Mitcha. 1997. "A study of USAF air traffic controller shiftwork: Sleep, fatigue, activity, and mood analyses." *Aviation, Space, and Environmental Medicine* 68: 18-23.

Miglioretti, D. L., C. C. Gard, P. A. Carney, T. L. Onega, D. S. Buist, E. A. Sickles, K. Kerlikowske, R. D. Rosenberg, B. C. Yankaskas, B. M. Geller, and J. G. Elmore. 2009. "When radiologists perform best: the learning curve in screening mammogram interpretation." *Radiology* 253 (3): 632-640. https://doi.org/10.1148/radiol.2533090070.

National Research Council. 2015. *Improving Diagnosis in Health Care*. Washington, D.C.: National Academies Press.

Nodine, C. F., C. Mello-Thoms, H. L. Kundel, and S. P. Weinstein. 2002. "Time course of perception and decision making during mammographic interpretation." *AJR Am J Roentgenol* 179 (4): 917-923. https://doi.org/10.2214/ajr.179.4.1790917.

Nodine, CF, HL Kundel, C Mello-Thoms, SP Seinstein, SG Orel, DC Sullivan, and EF Conant. 1999. *Academic Radiology*. Vol. 6. 575 – 585.

Nodine, CG , and C. Mello-Thoms. 2019. "Acquiring expertise in radiologic image interpretation." In *The Handbook of Medical Image Perception and Techniques*, edited by E. Samei and EA Krupinski, 167 – 187. Cambridge, UK: Cambridge University Press.

Norman, G. R., S. D. Monteiro, J. Sherbino, J. S. Ilgen, H. G. Schmidt, and S. Mamede. 2017. "The Causes of Errors in Clinical Reasoning: Cognitive Biases, Knowledge Deficits, and Dual Process Thinking." *Acad Med* 92 (1): 23-30. https://doi.org/10.1097/ACM.0000000000001421.

Nuclear Regulatory Commission. 2013. *Shearon Harris Nuclear Plant – NRC Special Inspection.* Report 05000400/2012010. ML13192A154.

Nuclear Regulatory Commission. 2019. *Nuclear Regulatory Commission Vendor Inspection Report of Framatome Inc.* No. 99901300/2019-201. ML19261A188.

Pashler, H., J. C. Johnston, and E. Ruthruff. 2001. "Attention and performance." *Annu Rev Psychol* 52: 629-651. https://doi.org/10.1146/annurev.psych.52.1.629.

Poulton, E. C. 1973. "The effect of fatigue upon inspection work." *Applied Ergonomics* 4 (2): 73-83. https://doi.org/10.1016/0003-6870(73)90080-x.

Pugel, A. E., V. V. Simianu, D. R. Flum, and E. Patchen Dellinger. 2015. "Use of the surgical safety checklist to improve communication and reduce complications." *J Infect Public Health* 8 (3): 219-225. https://doi.org/10.1016/j.jiph.2015.01.001.

Regan, M. A., C. Hallett, and C. P. Gordon. 2011. "Driver distraction and driver inattention: definition, relationship and taxonomy." *Accid Anal Prev* 43 (5): 1771-1781. https://doi.org/10.1016/j.aap.2011.04.008.

Sanquist, TF, S Morrow, A D'Agostino, N Hughes, and C. Franklin. 2018. *Human Factors in Nondestructive Examination: Manual Ultrasonic Testing Task Analysis and Field Research.* PNNL-27441. NRC Adams ML18176A055.

Sheridan, H., and E. M. Reingold. 2017. "The Holistic Processing Account of Visual Expertise in Medical Image Perception: A Review." *Front Psychol* 8: 1620. https://doi.org/10.3389/fpsyg.2017.01620.

Singh, H., A. N. Meyer, and E. J. Thomas. 2014. "The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations." *BMJ Qual Saf* 23 (9): 727-731. https://doi.org/10.1136/bmjqs-2013-002627.

Smith-Coggins, R., S. K. Howard, D. T. Mac, C. Wang, S. Kwan, M. R. Rosekind, Y. Sowb, R. Balise, J. Levis, and D. M. Gaba. 2006. "Improving alertness and performance in emergency department physicians and nurses: the use of planned naps." *Ann Emerg Med* 48 (5): 596-604, 604 e591-593. https://doi.org/10.1016/j.annemergmed.2006.02.005.

Stephens, H. 2000. "NDE Reliability - Human Factors - Basic Considerations." Proceedings of the 15th World Conference on Non-Destructive Testing. https://www.ndt.net/article/wcndt00/papers/idn736/idn736.htm.

Strayer, DL , JM Watson, and FA Drews. 2011. "Cognitive Distraction While Multitasking in the Automobile." In *The Psychology of Learning and Motivation*, edited by Brian Ross, 29-58. Burlington Academic Press.

Tolsgaard, M. G., C. Ringsted, E. Dreisler, L. N. Norgaard, J. H. Petersen, M. E. Madsen, N. L. Freiesleben, J. L. Sorensen, and A. Tabor. 2015. "Sustained effect of simulation-based ultrasound training on clinical performance: a randomized trial." *Ultrasound Obstet Gynecol* 46 (3): 312-318. https://doi.org/10.1002/uog.14780.

Van Dongen, H. P., G. Maislin, J. M. Mullington, and D. F. Dinges. 2003. "The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation." *Sleep* 26 (2): 117-126. https://doi.org/10.1093/sleep/26.2.117.

Van Dongen, H. P., K. M. Vitellaro, and D. F. Dinges. 2005. "Individual differences in adult human sleep and wakefulness: Leitmotif for a research agenda." *Sleep* 28 (4): 479-496. https://doi.org/10.1093/sleep/28.4.479.

van Geel, K., E. M. Kok, J. Dijkstra, S. G. Robben, and J. J. van Merrienboer. 2017. "Teaching Systematic Viewing to Final-Year Medical Students Improves Systematicity but Not Coverage or Detection of Radiologic Abnormalities." *J Am Coll Radiol* 14 (2): 235-241. https://doi.org/10.1016/j.jacr.2016.10.001.

Waite, S., A. Grigorian, R. G. Alexander, S. L. Macknik, M. Carrasco, D. J. Heeger, and S. Martinez-Conde. 2019. "Analysis of Perceptual Expertise in Radiology - Current Knowledge and a New Perspective." *Front Hum Neurosci* 13: 213. https://doi.org/10.3389/fnhum.2019.00213.

Waite, S., J. Scott, B. Gale, T. Fuchs, S. Kolla, and D. Reede. 2017. "Interpretive Error in Radiology." *AJR Am J Roentgenol* 208 (4): 739-749. https://doi.org/10.2214/AJR.16.16963.

Weinger, M. B., and S. Ancoli-Israel. 2002. "Sleep deprivation and clinical performance." *JAMA* 287 (8): 955-957. https://doi.org/10.1001/jama.287.8.955.

Williamson, A., D. A. Lombardi, S. Folkard, J. Stutts, T. K. Courtney, and J. L. Connor. 2011. "The link between fatigue and safety." *Accid Anal Prev* 43 (2): 498-515. https://doi.org/10.1016/j.aap.2009.11.011.

Wynn, R. M., J. L. Howe, L. C. Kelahan, A. Fong, R. W. Filice, and R. M. Ratwani. 2018. "The Impact of Interruptions on Chest Radiograph Interpretation: Effects on Reading Time and Accuracy." *Acad Radiol* 25 (12): 1515-1520. https://doi.org/10.1016/j.acra.2018.03.016.

Yu, J. P., A. P. Kansagra, and J. Mongan. 2014. "The radiologist's workflow environment: evaluation of disruptors and potential implications." *J Am Coll Radiol* 11 (6): 589-593. https://doi.org/10.1016/j.jacr.2013.12.026.

Appendix A – Structured Interview Protocol

Encoded UT Examination Functions and Tasks

Supporting Material for Structured Interview

Pacific Northwest National Laboratory

June 26, 2020

Background

Thank you for agreeing to participate in a teleconference interview. This project is part of a program of research in the human factors aspects of NDE that Pacific Northwest National Laboratory (PNNL) is performing with the Nuclear Regulatory Commission (Carol Nove, project manager).

The research team has previously conducted similar interviews with examiners addressing manual conventional ultrasonic examinations, and we are now extending that work to address human performance and reliability with encoded ultrasonic exams.

We will use material in the attached pages to structure our discussion with you. Please review this material briefly before our scheduled teleconference – which we expect will take about an hour.

The interviews are confidential, and you will not be identified in association with your responses.

If you have any questions before our scheduled call, please contact either:

Joel Harrison, Joel.Harrison@pnnl.gov

(509) 375-4504

Tom Sanquist, Sanquist@pnnl.gov

(206) 528-3240

Interview Questions

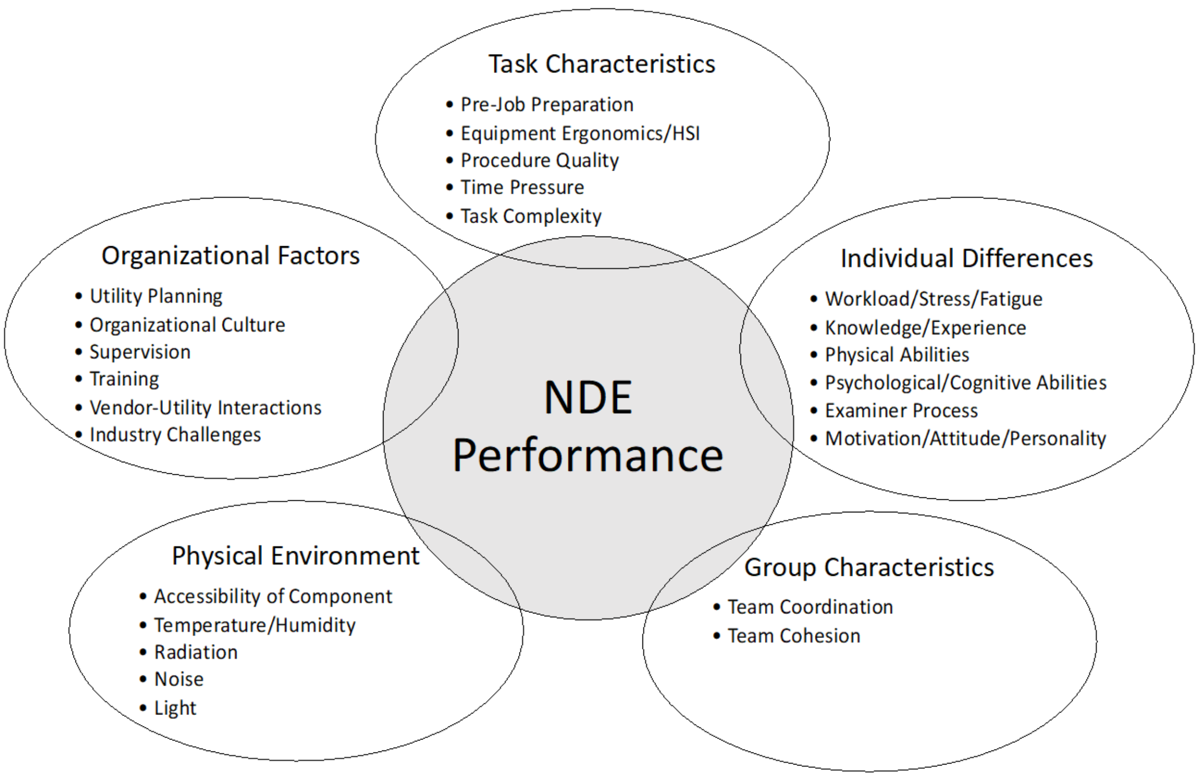
Function/Task Specific Questions (Refer to pages 6-11 for each function and associated tasks)

- As an examiner, what is your role in this function?
- What tasks are most important to get right? Why?
- Are there any tasks specific to encoded UT not on this list that you think are important?
- Does performing this function differ in any substantial way from when it is performed for manual conventional UT (other than the tasks shown)?
    - For encoded UT, is the potential for errors reduced, increased, or does it stay the same? Expand as necessary
- What is the role of experience in performing this function?

Overall Questions

- What factors influence how well encoded UT tasks are performed? What kinds of problems are encountered if the tasks are not performed well? (Refer to the graphic and table for Performance Influencing Factor definitions)
- Are there heuristics – "rules of thumb" -- that are used when doing encoded UT? Have these ever proved wrong or not applicable?
- Are there times when you need to adapt or change what you had intended to do?
- Are there common error traps that examiners can fall into when performing encoded UT?
- Are there any specific incidents you recall when an encoded UT exam resulted in error or a near-miss situation? What were the circumstances?

Performance Influencing Factors

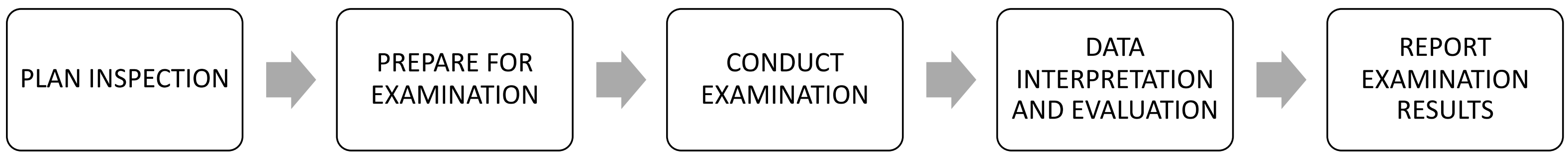


4

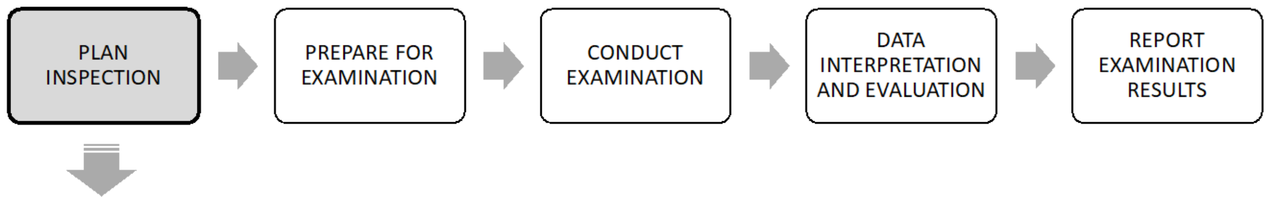
Category	Performance Influencing Factors	Definition
Task Characteristics	Pre-job Preparation	Activities performed by the vendor to get ready for an exam
	Equipment	Specific equipment and material used to perform exam, and its availability
	Procedure	Content and/or nature of a written procedure used in an exam
	Time Pressure	Temporal constraints due to specific exam performance or the overall inspection schedule
	Task Complexity	Factors such as ambiguity in assessing or executing the task, the degree of mental effort or knowledge involved, whether special sequencing or coordination is required, or whether the task requires sensitive and careful manipulations
Individual Differences	Knowledge/Experience	What the examiner knows, level of experience on the job, and certifications and qualifications
	Examiner Process	How the examiner executes the specific task (reporting comments, etc.)
	Motivation/Attitude/Personality	Characteristics of the person
	Physical Abilities	Height, weight, dexterity, etc.
	Cognitive Factors	Attention, perception, memory, spatial ability
	Workload/Stress/Fatigue	Pace, intensity, and duration of exam, work shifts or assignments
Group Characteristics	Team Coordination	Peer interaction while doing the inspection. This should be focused on performing the exam.
	Team Cohesion	Familiarity of inspection team with one another and the impacts upon exam process
Physical Environment	Accessibility of Component	Location, reachability
	Lighting	Visibility
	Noise	Ability to hear while doing exam
	Radiation	Task-specific dose, cumulative dose
	Temperature and Humidity	Heat, dehydration, glasses fogging, etc.
Organizational Factors	Utility Planning	Activities conducted by the utility to prepare for exam
	Organizational Culture	Norms and expectations in the work environment (the "feeling of the workplace")
	Supervision	Oversight of the NDE process—either directly by vendor, utility, or regulator
	Training	Comments about types and quality of training, required and optional practice on samples
	Vendor-Utility Interactions	Working relationships between the two parties, including developing work packages, documentation requirements, expectations for reporting indications, role played by vendor (sometimes as in-house NDE planner), etc.
	Industry Challenges	Work force availability, work opportunities

5

Encoded UT Functions



FUNCTIONS



TASKS

1. Develop plan based on requirements and previous data
2. Identify qualified examiners
3. Ensure component prepared for examination in the field

7

FUNCTIONS

PLAN INSPECTION → **PREPARE FOR EXAMINATION** → CONDUCT EXAMINATION → DATA INTERPRETATION AND EVALUATION → REPORT EXAMINATION RESULTS

TASKS

1. Receive plant orientation and training
2. Review work package and procedure
   a) Develop scan plan
3. Verify that component is within the procedural and personnel qualification ranges
4. Assemble materials and equipment.
   a) Test equipment functionality prior to staging.
   b) Verify data software dataset.
5. Calibrate equipment to procedure
6. Attend pre-job briefing
7. Coordinate examiner responsibilities
8. Pack equipment and materials
9. Prepare to enter area for exam
10. Locate component

8

FUNCTIONS

PLAN INSPECTION → PREPARE FOR EXAMINATION → **CONDUCT EXAMINATION** → DATA INTERPRETATION AND EVALUATION → REPORT EXAMINATION RESULTS

TASKS

1. Verify at correct component and weld
2. Verify conditions at weld
3. Perform supplemental inspection measurements
4. Set up for scan
5. Perform calibration check
6. Mount scanner drive mechanism
   a) Perform pre-scan verification
7. Perform scan
   a) Monitor signal characteristics; adjust settings as appropriate
8. Take notes and pictures as required
9. Verify data complete
   a) Data analyst ensures coverage of Code required volume and data quality
10. Dismount drive mechanism and clean up

9

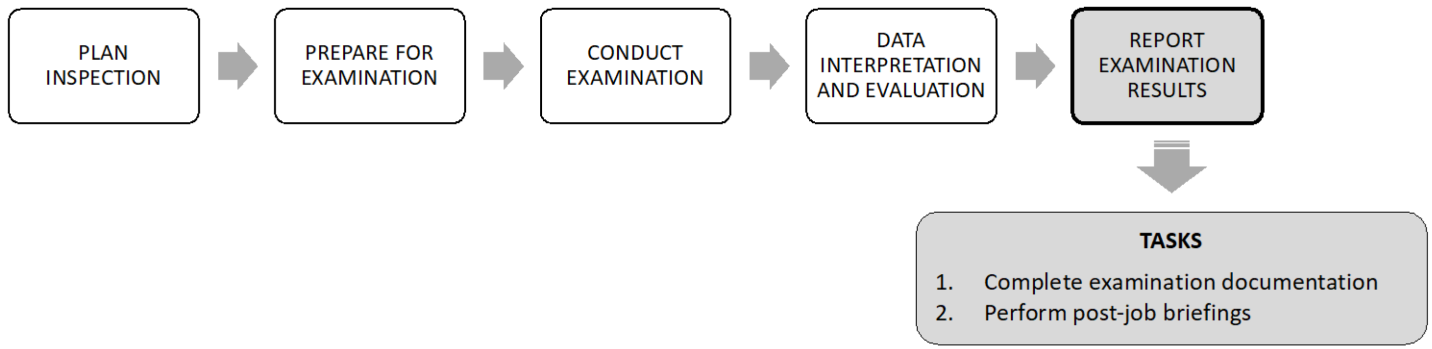
FUNCTIONS

PLAN INSPECTION → PREPARE FOR EXAMINATION → CONDUCT EXAMINATION → **DATA INTERPRETATION AND EVALUATION** → REPORT EXAMINATION RESULTS

TASKS

1. Verify post-examination calibration
2. Flaw detection analysis
   a) Gate volume to identify benchmark information
   b) Gate regions within volume; compare patterns and responses from different wave modes, angles, skews and beam directions
   c) Adjust sensitivity (gain or palette) to improve contrast as necessary
3. Classify indications as flaw or geometry
4. Determine size and depth per procedure
5. Record non-relevant indications in sufficient detail to assist future examinations

FUNCTIONS



11

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)

www.pnnl.gov | www.nrc.gov