
Evaluation of the Computerized Procedures Manual II (COPMA II)

Manuscript Completed: May 1995
Date Published: November 1995

Prepared by
S. A. Converse

North Carolina State University
Raleigh, NC 27695-7801

J. Persensky, NRC Project Manager

Prepared for
Division of Systems Technology
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001
NRC Job Code L1979

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED *SR*

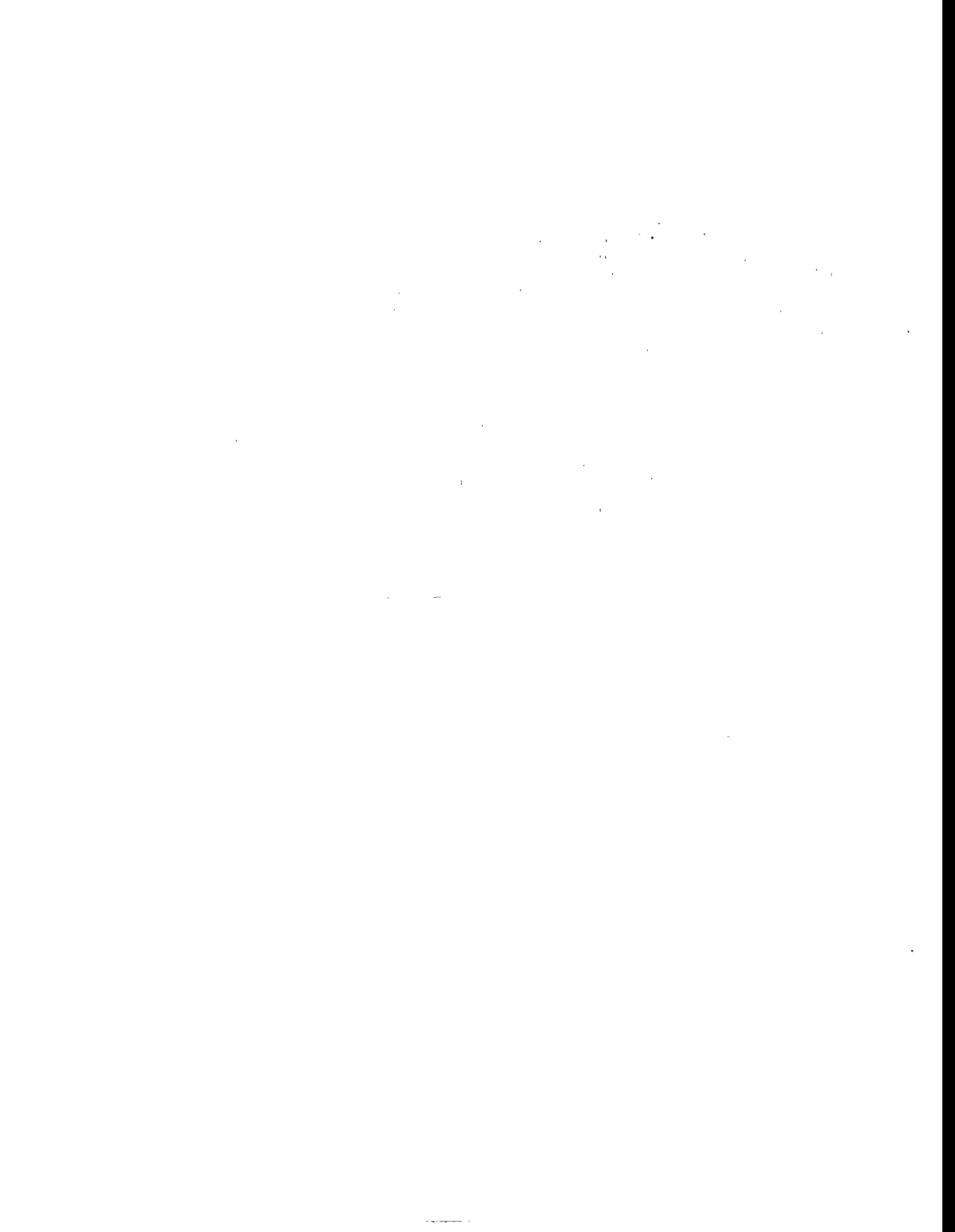
MASTER

DISCLAIMER

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

ABSTRACT

The purpose of this study was to evaluate the effects of a computerized procedure system, the Computerized Procedure Manual II (COPMA-II), on the performance and mental workload of licensed reactor operators. To evaluate COPMA-II, eight teams of two operators were trained to operate a scaled pressurized water reactor facility (SPWRF) with traditional paper procedures and with COPMA-II. Following training, each team operated the SPWRF under normal operating conditions with both paper procedures and COPMA-II. The teams then performed one of two accident scenarios with paper procedures, but performed the remaining accident scenario with COPMA-II. Performance measures and subjective estimates of mental workload were recorded for each performance trial. The most important finding of the study was that the operators committed only half as many errors during the accident scenarios with COPMA-II as they committed with paper procedures. However, time to initiate a procedure was fastest for paper procedures for accident scenario trials. For performance under normal operating conditions, there was no difference in time to initiate or to complete a procedure, or in the number of errors committed with paper procedures and with COPMA-II. There were no consistent differences in the mental workload ratings operators recorded for trials with paper procedures and COPMA-II.



CONTENTS

	<u>Page</u>
ABSTRACT	iii
EXECUTIVE SUMMARY	ix
ACKNOWLEDGEMENTS	xii
1 INTRODUCTION	1
1.1 Background	1
1.2 COPMA-II	1
1.2.1 Procedure Editor-II	2
1.2.2 COPMA-II On-Line System	2
1.2.3 Uses and Functions of COPMA-II	5
2 EVALUATION OF COPMA-II	6
2.1 Study Objectives	6
2.2 Scaled Pressurized Water Reactor Facility	6
2.3 Subjects	8
2.4 Experimental Task	8
2.5 Independent Variables	8
2.5.1 Task Type Variable	8
2.5.2 Procedure Type Variable	9
2.5.3 Team Variable	10
2.5.4 Trial Block Variable	10
2.5.5 Control for Order Effects	10
2.6 Dependent Variables	10
2.6.1 Response Time Measures	12
2.6.2 Error Measures	12
2.6.3 Subjective Measures	12
2.7 Research Procedures	13
2.7.1 SPWRF Training	13
2.7.2 Research Setting for Training and Data Collection	14
2.7.3 Performance of the Change of Power and Accident Scenario Tasks	15
2.8. Data Sets and Study Designs	15
2.8.1 Performance Measures	15
2.8.2 Subjective Measures	15
2.9 Data Analysis	16

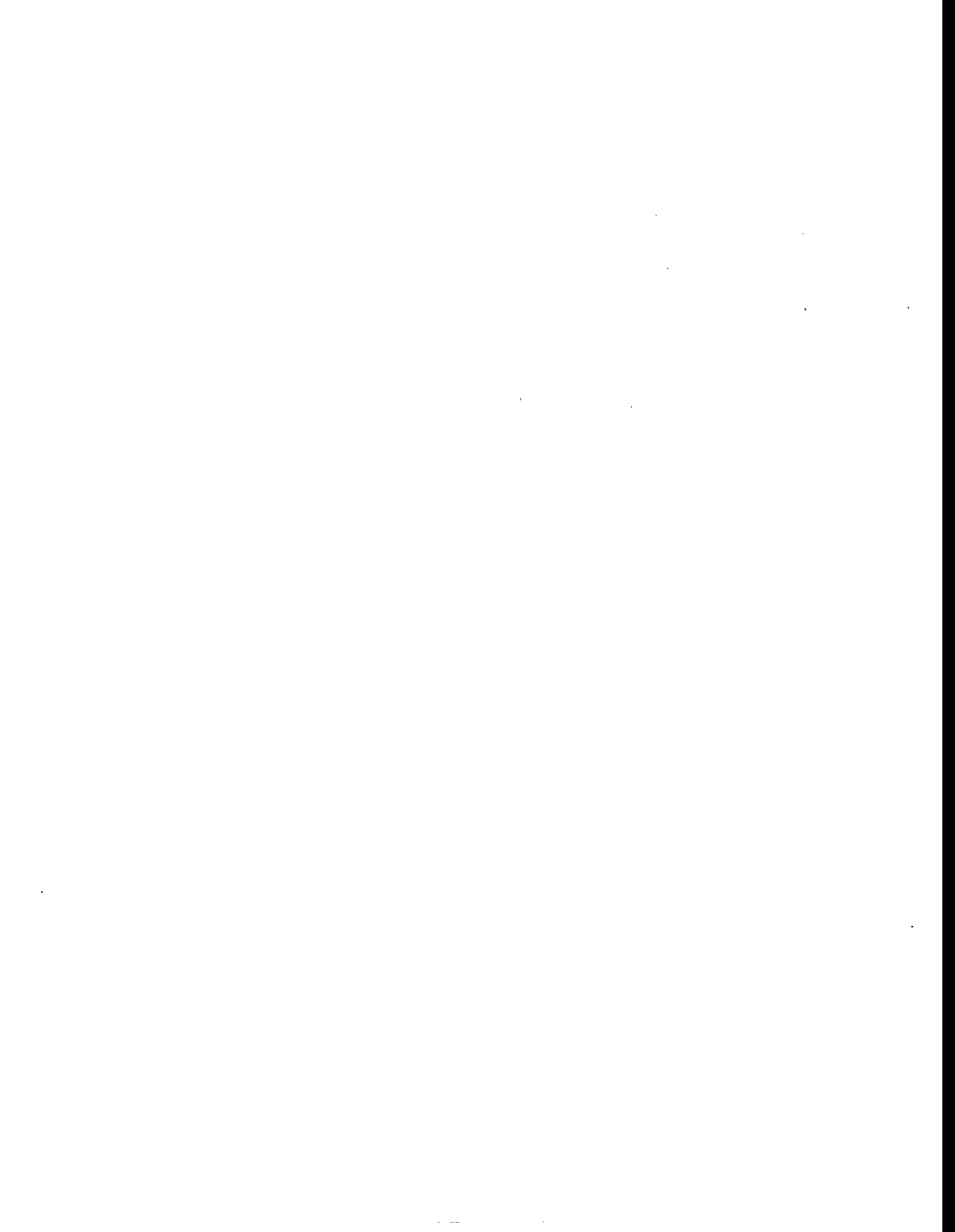
	<u>Page</u>
3 RESULTS	17
3.1 Summary of Results	17
3.2 Performance Measures	17
3.2.1 Effects of Procedure Type	17
3.2.2 Effects of Trial Block	19
3.3 Subjective Workload Ratings	19
3.3.1 Temporal Demand	19
3.3.2 Self-Ratings of Performance	19
3.4 Performance Differences Between Teams	19
3.5 Differences in Subjective Ratings Between Teams	21
3.5.1 Effect of Procedure Type on Individual Teams	21
3.5.2. Other Team Differences in Subjective Ratings of Workload	23
4 DISCUSSION OF STUDY RESULTS	25
4.1 Study Objectives and Hypotheses	25
4.2 Accident Scenario Data	25
4.2.1 Error Data	25
4.2.2. Response Time Data	26
4.3 Subjective Workload Data	27
4.4 Research Methods and Lessons Learned	28
4.4.1 Effects of Task Type	28
4.4.2. Performance Measures	29
4.4.3. Effects of Practice	29
4.4.4. Differences in the Skill Level of Operators	29
5 REFERENCES	31
Appendix A: Timing Measures Recorded During the Change of Power Tasks	A-1
Appendix B: Timing Measures Recorded During the Loss of Coolant Accident Scenario	B-1
Appendix C: Timing Measures Recorded During the Steam Generator Tube Rupture Accident Scenario	C-1
Appendix D: NASA-TLX Subjective Workload Scale: Instructions Read to Subjects and Prototype Scale	D-1

FIGURES

1.1	The PED-II Screen Layout	3
1.2	COPMA-II On-Line System Interface	4
2.1	Interior of the Scaled Pressurized Water Reactor Facility	7
3.1	Accident Scenario Errors by Procedure Type	18
3.2	RTC for Accident Scenario by Procedure Type	18
3.3	Initial Change of Power Task RTI for Procedure Type by Team	20
3.4	Initial Change of Power Task RTC for Procedure Type by Team	20
3.5	Accident Scenario RTC for Procedure Type by Team	21
3.6	Accident Scenario Errors for Procedure Type by Team	22
3.7	NASA-TLX Ratings of Mental Demand for Initial Change of Power Task for Procedure Type by Team	22
3.8	Initial Change of Power Task NASA-TLX Ratings of Frustration for Procedure by Team	23
3.9	Experimental Change of Power Task NASA-TLX Ratings of Frustration by Team	24

TABLES

2.1	Type of Procedure Used First by Teams for the Change of Power Tasks	11
2.2	Type of Accident Scenario and Procedure Type Used First by Teams	11



EXECUTIVE SUMMARY

This report concerns a study that was conducted to evaluate the effects of a computerized procedure system, the Computerized Procedure Manual-II (COPMA-II), on the performance and perceived mental workload of licensed reactor operators (ROs). COPMA-II contains advanced features that display system parameters, and provides graphics that allow operators to track their progress through several parallel sets of procedures. COPMA-II also provides automatic monitoring and feedback of prespecified system parameters.

COPMA-II incorporates two main systems, the Procedure Editor II (PED-II) and the COPMA-II on-line system. The PED-II is used to create textual and graphical representations of procedures that are entered into the data base of the COPMA-II on-line system. The textual representations serve as the mechanism by which operators locate and follow procedures, while the graphical representations are intended to help operators visualize the relationships between procedure steps, and determine how far they have advanced in one or more procedures.

The COPMA-II on-line system is the component of COPMA-II that is used by ROs to select, execute, and monitor procedures in the control room. The interface of the on-line systems contains five panes. The Main Menu Pane is where operators activate procedures and view values of automatically monitored system parameters. The Bookshelf Pane displays a list of all procedures that are contained in the COPMA-II data base. The Instruction Pane displays previously executed, current, and yet to be executed procedural instructions. Finally, the Flowchart Pane provides a graphical overview of procedure instructions that have been activated by the RO.

The effects of COPMA-II on the performance and perceived mental workload of ROs was evaluated in a study that was conducted at North Carolina State University. The study employed eight licensed reactor operators and eight licensed senior reactor operators (SROs) who controlled the Scaled Pressurized Water Research Facility (SPWRF) in the Department of Nuclear Engineering. The 16 operators were divided into eight teams of two operators, with each team containing one RO and one SRO.

The operators were trained and tested at the SPWRF over the course of two days. Training consisted of four one-hour sessions. The first session employed a lecture/demonstration format in which the operators were familiarized with the thermodynamics and function of the SPWRF, and with the SPWRF control room interface. During the second training session, ROs and SROs practiced operating the SPWRF with paper procedures. The third training session was a lecture/demonstration in which the operators were familiarized with the structure of COPMA-II. During the fourth training session, operators practiced operating the SPWRF with COPMA-II.

Operators in each team performed two initial change of power (COP) tasks, two experimental COP tasks, and two accident scenarios. The initial COP tasks were performed at the end of training to confirm that operators in each team were equally proficient at operating the SPWRF. Each team performed one initial COP task maneuver with paper procedures, and the remaining COP task

maneuver with COPMA-II. The experimental COP tasks were conducted at the beginning of the data collection sections to simulate performance under normal operating conditions. One experimental COP task was performed with paper procedures, while the remaining experimental COP task was performed with COPMA-II. After the teams had performed the COP tasks, performance under accident conditions was assessed by requiring operators to perform two simulated accident scenarios, a small break loss of cooling accident (LOCA) and a steam generator tube rupture (SGTR). Each team performed one accident scenario with paper procedures and the remaining scenario with COPMA-II.

The performance measures collected for each team during each data collection session included: (1) time to initiate a response (RTI), time to complete a response (RTC); and the number of errors committed. Subjective estimates of mental workload were collected by having operators complete the NASA-TLX Mental Workload Scale after each performance trial.

The study data were submitted to one of four Multivariate Analyses of Variance (MANOVA) tests. One MANOVA was used to evaluate the effects of procedure type (paper procedures vs. COPMA-II), task type (initial COP, experimental COP, or accident scenario tasks), and team (Team 1 - Team 8) on COP task performance measures. A second MANOVA was used to test the effects of the same factors on the accident scenario data. The third and fourth MANOVAs were used to evaluate the effects of the above mentioned factors, plus the effects of operator type (RO vs. SRO), on subjective measures for the COP tasks and for the accident scenario tasks, respectively.

The most important finding of the study was that the operators committed only about half as many accident scenario errors with COPMA-II as they committed with paper procedures. The COPMA-II error advantage for the accident scenarios is especially impressive because it was obtained when operators had only a relatively small amount of practice with COPMA-II.

While our study does not provide conclusive evidence about why relatively few accident scenario errors were made with COPMA-II, the comments of the operators do provide some clues. Several operators commented that COPMA-II did not allow them to "skip ahead" and preview future procedure steps, as they could do easily with paper procedures. Because we defined an error as any deviation from a predefined optimal sequence of procedure steps, temporarily leaving the predefined procedure path to look ahead constituted an action that was counted as an error in this study. If operators did look ahead when using paper procedures, but were forced to be more methodical when using COPMA-II, the error rate would be highest for paper procedures.

While COPMA-II markedly reduced the number of accident scenario errors, RTI for the accident scenarios was slower for COPMA-II than for paper procedures. Thus, taken together, the accident scenario error and RTI data describe a speed-accuracy tradeoff in which COPMA-II reduced the number of errors, but lengthened the time required to initiate responses.

The speed-accuracy trade-off engendered by COPMA-II in the accident scenarios probably does not indicate that COPMA-II would seriously degrade operator performance in operational conditions. Errors are usually a more critical factor in recovery from accidents than is the speed with which emergency operating procedures are initiated. Thus, the finding that there were the least errors committed with COPMA-II in the accident scenarios is probably a

more meaningful finding than the finding that accident scenario RTI was slower for COPMA-II than for paper procedures.

The findings of this study provide some information about the research methods that should be used to validate devices such as computerized operating procedures. Four main methodological issues are discussed. First, the types of tasks that should be used to validate devices such as computerized procedures are considered. Second, the selection of performance measures is addressed. Third, the amount of time that operators should be allowed to practice with new devices before validation data are collected is discussed. Finally, the effect of differences in skill level between operators is considered.

ACKNOWLEDGEMENTS

Several individuals were instrumental in the successful completion of this project. First, I would especially like to thank Eyvind Ness and John Tiegen, as well as other professionals who work for the OECD Halden Reactor project. These individuals gave freely of their time and expertise, and were ready to help during all hours of the day and night. The project could not have been finished without them. I would also like to thank Eric Wiebe for sharing his expertise with the UNIX operating system on several occasions. Finally, my thanks go to Dr. J. J. Persensky, my technical monitor at the Nuclear Regulatory Commission. His assistance and good sense of humor were always appreciated.

1 INTRODUCTION

1.1 Background

The role that human error plays in events and accidents in nuclear power plants (NPPs) has long been an important concern for industry professionals [1]. Analysis of event reports shows that, while many factors contribute to unsafe conditions, the most commonly cited cause of events in the United States is operating procedures. In fact, procedural problems have been cited as a contributing event in approximately 69% of event reports [2] [3]. As a result, the U.S. Nuclear Regulatory Commission initiated a programmatic effort to evaluate and encourage industry to improve emergency operating procedures [4].

Because components of nuclear power plants are closely coupled (e.g., pressurizer temperature and pressure), the initiation of a single procedure may require operators to perform several actions simultaneously. In addition, the abnormality of one parameter will affect other parameters as well. Consequently, operators are sometimes required to use several EOPs simultaneously. When using multiple procedures, operators are sometimes required to interrupt the performance of a unified sequence of procedural steps to perform a related but subordinate set of actions [4] [5]. This means that operators must switch back and forth between procedures, keeping track of their place in each procedure.

The problems associated with switching between several EOPs is exacerbated by the paper medium that is typically used to present EOPs. Operators often report that it is difficult and time-consuming to locate the correct EOP within the manual that is used to hold them [6]. Another common complaint is that it is difficult to move back and forth between pages of the EOP without losing track of which steps have been completed. If operators try to solve this problem by removing relevant EOPs from the manual, the pages may become mixed up or get lost. One potential solution to these problems is to adopt a flowchart format for the EOPs. However, while flowcharts solve manual problems associated with paper procedures, they don't resolve the problem of keeping one's place in the EOPs when multiple procedures and multiple flowcharts are used.

Computerized procedures have been suggested as a mechanism for aiding operators by making it easier to locate procedures, to track one's exact place in each active procedure, and to move back and forth between procedures by allowing operators to quickly identify the specific return point in a latent procedure when a branching procedure step has been completed [5] [6].

1.2 COPMA-II

Since 1985, researchers at the OECD Halden Reactor Project in Halden, Norway have explored the feasibility of converting paper procedures to a computerized system. They have attempted to identify the advantages of a computerized procedure format (e.g., error reduction and increased response speed), and have analyzed the ability of operators to work effectively with computerized procedures [7].

To evaluate the effectiveness of computerized procedures, members of the OECD Halden Reactor Project created an initial prototype of a computerized procedure system which they called the Computerized Procedure Manual (COPMA) [5]. The original COPMA system was evaluated and tested in the Halden Project's advanced experimental control room facility, the Halden Man-Machine Laboratory. The results of this research were then used to create a second version of COPMA, COPMA-II [8]. COPMA-II is intended to reflect a substantial enhancement of the ease-of-use provided of the original version of COPMA.

1.2.1 Procedure Editor-II

COPMA-II incorporates two main systems, the Procedure Editor II (PED-II) and the COPMA-II on-line system. The PED-II is the COPMA-II editor that is used to create textual and graphical representations of procedures that will be entered into the data base of the COPMA-II on-line system. The PED-II is illustrated in Figure 1.1. It consists of a menu window and a graphical window. The menu window provides a list of commands that are used to create a textual representation of the steps contained in each procedure. Each textual representation describes a set of steps that must be performed to complete each instruction contained in a specific procedure. As the PED-II is commanded to create a textual representation, it automatically constructs a graphical representation as well. This image is depicted on PED II's graphical window as a flow-chart that depicts procedural steps in a tree-like structure.

The textual and graphical representations of procedure steps will eventually be entered into the data base of the COPMA-II on-line system. The textual representations will serve as the mechanism by which operators locate and follow procedures, while the graphical representations are intended to help operators visualize the relationship between procedure steps, and determine how far they have advanced in each of several procedures [9].

1.2.2 COPMA-II On-Line System

The on-line system is the component of COPMA-II that is used by reactor operators to select, execute, and monitor procedures in the control room [9]. As can be seen in Figure 1.2, the interface of the on-line COPMA-II system consists of five panes: (1) Main Menu Pane; (2) Bookshelf; (3) Desk Pane; (4) Instruction Pane; and (5) Flowchart Pane [9].

The Main Menu Pane is positioned at the top left of the on-line system interface. Operators use buttons displayed on this pane to activate procedures, instruct COPMA-II to monitor parameters automatically, view the values of automatically monitored parameters, or to request system or help information.

The Bookshelf Pane displays a list of all procedures that are contained in the COPMA-II data base.

The Desk Pane provides an overview of all procedures that are active (available for execution) at the present time. Two buttons at the bottom of this pane allow operators to request automatic monitoring of specific parameters, and to request a list of all parameters that are currently being monitored by COPMA-II.

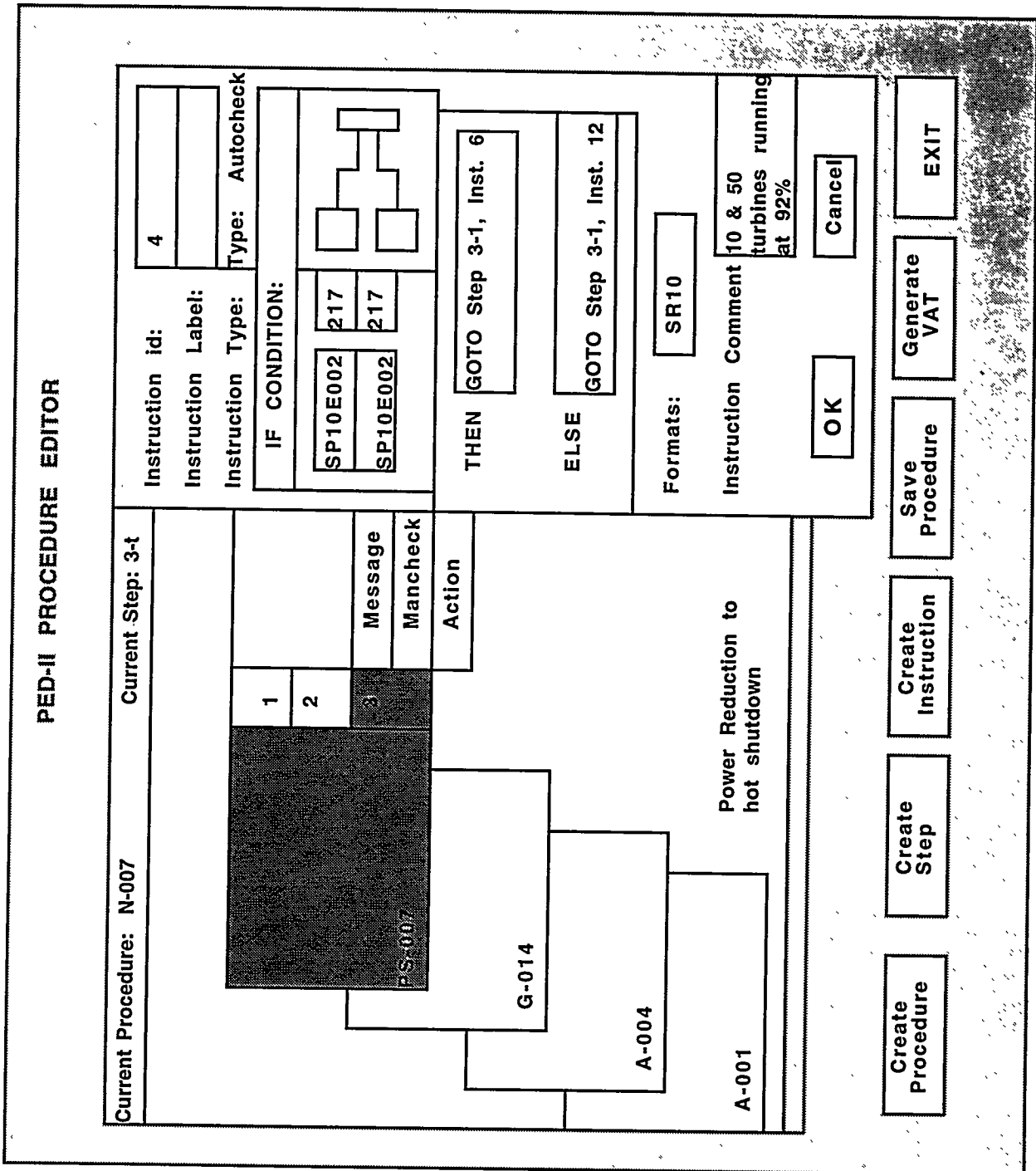


Figure 1.1 The PED-II Screen Layout

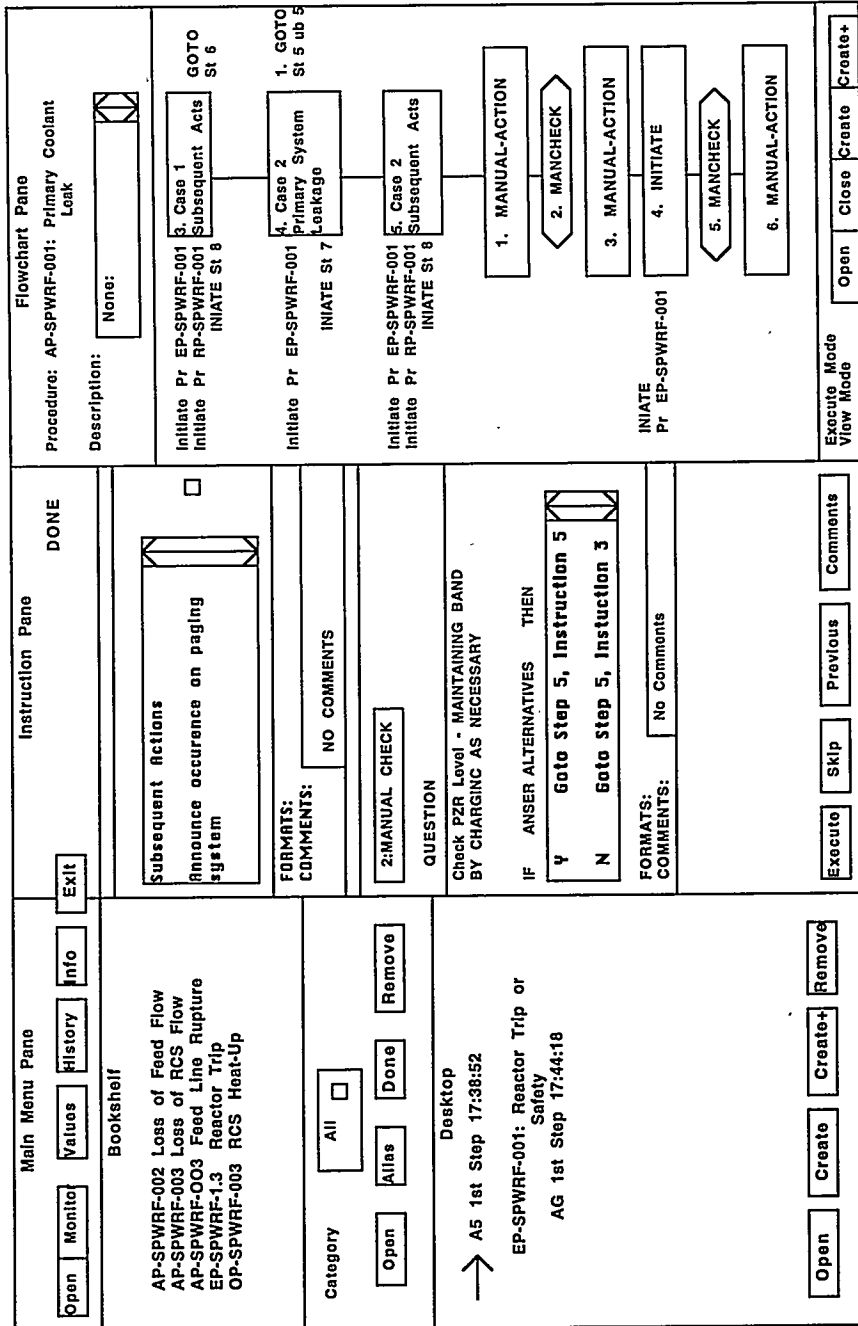


Figure 1.2 COPMA-II On-Line System Interface

In the middle of the COPMA-II interface is the Instruction Pane which displays the name, description, and type of each active procedure. The Instruction Pane is divided into three horizontal windows that display previously executed, current, and yet to be executed instructions from top to bottom, respectively. The text of the instructions is color coded in terms of a procedure step's current status; yellow text for previously executed instructions, red for current instructions, and blue for yet to be executed instructions. The four buttons at the bottom of the Instruction Pane allow operators to bypass the current instruction, return to a previously executed instruction, or execute the current procedure.

The Flowchart Pane provides the graphical overview of procedure instructions that was created by the PED-II. Due to the length of some procedures, the procedure graphic is presented in a scrollable text window. Each instruction is surrounded by a small box that is color coded in the same manner described for the Instruction Pane. The boxes are connected with vertical lines that illustrate the sequential relationships between the instructions [9]. The two buttons at the bottom of the Flowchart Pane allow operators to view, but not execute, procedures, and to return to the execute mode.

1.2.3 Uses and Functions of COPMA-II

COPMA-II can be used in two ways. When COPMA-II is used as an off-line system, it simply serves as a computerized mechanism for the storage and presentation of procedures. To execute procedure steps, the senior reactor operator (SRO) reads the instruction on the COPMA-II interface, and the reactor operator (RO) performs the specified actions manually. When used in this way, COPMA-II was designed to have several advantages over paper procedures. It provides more information per screen than is contained on a single page of paper procedures, allowing operators to scan procedures more rapidly than can be accomplished with paper procedures [6]. Thus, COPMA-II was also designed to present and organize information in an more useful fashion than is found in a paper manual's index or table of contents.

The second way in which the COPMA-II on-line system can be used is as an advanced operator aiding technique. The on-line system contains many high level functions that should support reactor operators [8]. For example, COPMA-II allows operators to enter control commands directly through the COPMA-II interface by pushing the execute button on the Instruction Pane. This COPMA-II function should allow operators to implement commands more quickly. [9]. The automatic monitoring feature of COPMA-II can be instructed to notify operators when a parameter has reached or exceeded a desired state. This feature should reduce operators' mental workload by reducing the number of parameters that must be checked periodically.

2 EVALUATION OF COPMA-II

Although COPMA-II was created to aid reactor operators, it must be validated before it can be used in an operational setting. When new systems are not validated, they may degrade, rather than enhance, performance. This is particularly true for computerized procedures because there are few guidelines available for their design, review, or implementation [4].

2.1 Study Objectives

To validate COPMA-II, a study with two specific objectives was performed. One objective was to determine if COPMA-II enhances or degrades the speed and accuracy of operators' performance under normal operating and accident conditions. To achieve this objective the speed and accuracy of operators' performance with paper procedures and with COPMA-II was compared. The second objective of the study was to compare operators' subjective ratings of mental demand, effort, frustration, temporal demand, self-performance, and physical demand for performance with paper procedures and with COPMA-II. This objective was achieved by recording operators' responses to the NASA-TLX Subjective Workload Scale.

2.2 Scaled Pressurizer Water Reactor Facility

The validation study was conducted at the Scaled Pressurized Water Reactor Facility (SPWRF) which is located in the Department of Nuclear Engineering at North Carolina State University. The control room of the SPWRF is depicted in Figure 2.1. The SPWRF is a 1/9-scale working model of the Prairie Island Nuclear Generating Plant's nuclear steam supply system, which is a two-loop Westinghouse Pressurized Water Reactor (PWR) [10]. The SPWRF uses two independent data acquisition and control computers to simulate the nuclear reactor core characteristics of a PWR, so the facility behaves like the prototype reactor. However, the SPWRF does not use nuclear fuel, but uses electric heaters to simulate the reactor core. Freon, rather than water, is employed as the working coolant to reduce the time and pressure required to heat the working fluid to a vapor state. As a result, the specific set points for the SPWRF are different from the set points used when water, rather than freon, is the working coolant.

The SPWRF distinguishes itself from most large scale thermal-hydraulic loops in that full operator interaction is possible for both steady-state and transient conditions, including normal operation, small break loss of coolant accidents (LOCAs), steam generator tube ruptures (SGTRs), and main steam line breaks. Safety features are designed into the system to prevent inadvertent over-pressurization of both the primary and secondary systems.

Almost every major component of a commercial PWR is contained in the SPWRF. Both primary and secondary sides of the plant are represented. The secondary side of the SPWRF is quite complete, containing condensers, condensate and feed pumps, feed water heaters, auxiliary feed water pumps, and turbine throttle valves [10].

The interface of the SPWRF is instrumented in such a way that all major operating parameters are continuously displayed on the control console. All control functions that affect plant response to the transients of interest are

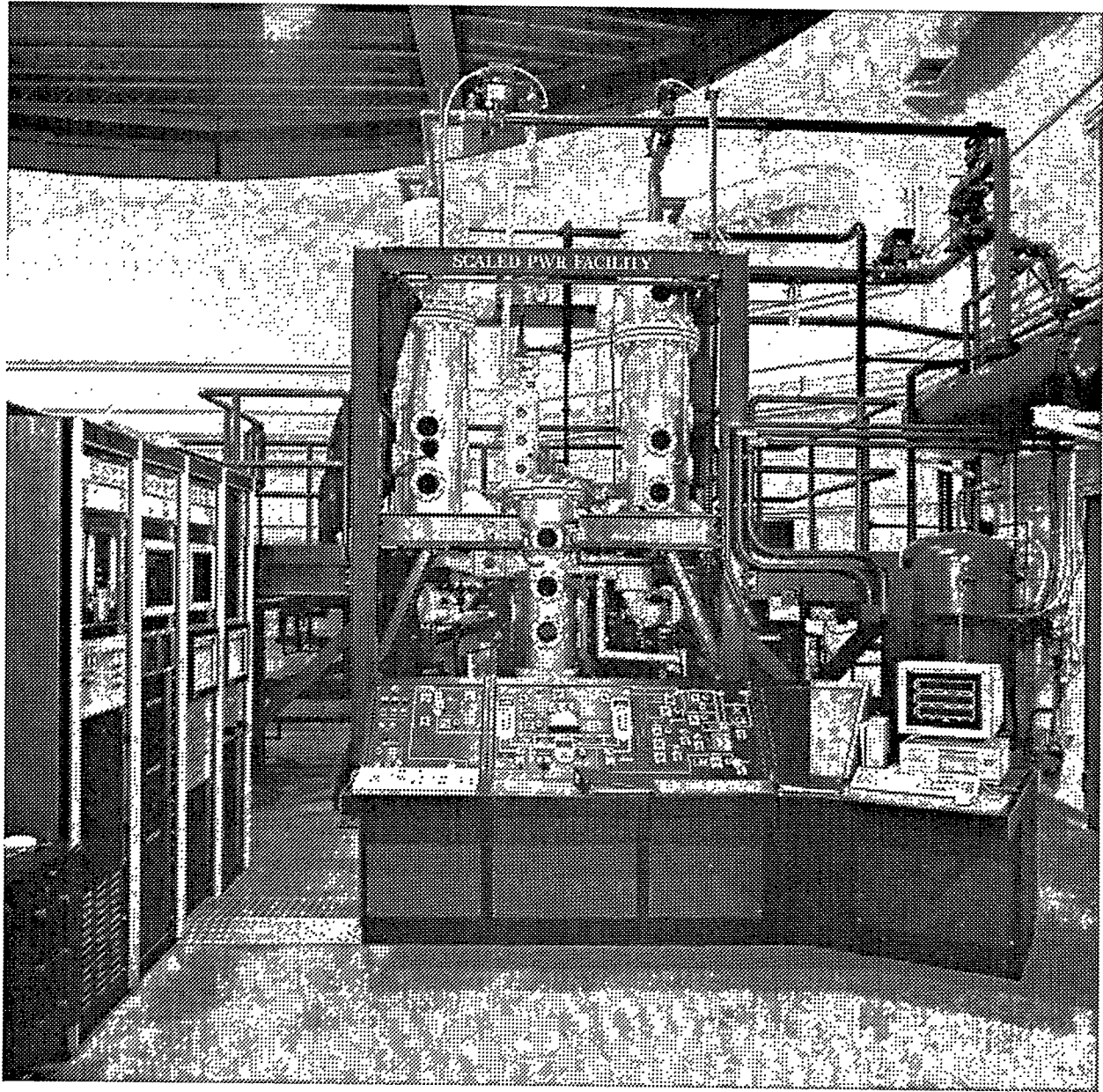


Figure 2.1 Interior of the Scaled Pressurized Water Reactor Facility

provided. Several computers are available for graphic display of system conditions in a variety of easily understood formats.

There are, of course, some differences between the SPWRF control room and a typical PWR control room. The SPWRF control room is approximately 1/3 as large as a typical PWR control room. Thus, the interface of the SPWRF is also less complex than the interface of a typical PWR control room. For example, there is less redundancy of instrument readings on the SPWRF interface than is typically the case for control room interfaces. There are also fewer annunciators. Although both analog and digital displays are provided, the SPWRF interface uses a relatively high number of digital displays. The SPWRF interface does not include an electrical generator turbine control panel because the SPWRF does not contain the electrical turbine side of a PWR.

2.3 Subjects

Sixteen licensed operators (8 reactor operators, 8 senior reactor operators) from a commercial power plant participated in the study. The 16 operators were divided into eight teams, with each team containing one senior reactor operator (SRO) and one reactor operator (RO). Each operator was a male volunteer who had between seven and 18 years of experience as a licensed operator. None of the operators had controlled the SPWRF for at least one year, but two operators had used the SPWRF during a one week training session in 1991.

Operators were recruited by posting a notice in the control rooms of a privately owned PWR facility. The notice described the study and asked SROs who were interested in participating to contact the researchers. When this contact was made, each SRO was asked to recruit a RO with whom he felt confident working to serve as his partner in the study.

2.4 Experimental Task

The experimental task required operators to control the SPWRF. Each SRO and RO performed the functions that they traditionally perform at their home plant. Thus, the SRO was responsible for locating and reading the procedures, for requesting diagnostic information, and for making required notification of events to administrative personnel via the telephone. The RO monitored plant parameters, provided information requested by the SRO, and manually controlled the SPWRF as specified by the SRO.

2.5 Independent Variables

The variables evaluated in this study were: (1) task type (change of power vs. accident scenario); (2) procedure type (COPMA-II vs. paper procedures); (3) team (teams 1-8); (4) trial block (1 vs. 2); and (5) operator type (RO vs. SRO).

2.5.1 Task Type Variable

Operators in each team performed an initial change of power (COP) task, an experimental COP task, and two accident scenarios. The initial COP task was performed at the end of the training sessions as a method of confirming that operators in each team were equally proficient at operating the SPWRF. The

experimental COP task was performed at the beginning of the data collection sections, and was followed by the performance of the two accident scenarios.

During the initial COP task, the operators performed four COP maneuvers; (1) power increase with paper procedures; (2) power decrease with paper procedures; (3) power increase with COPMA-II; and (4) power decrease with COPMA-II. A power increase was performed by changing reactor power from 40% to 70% (where 80 kw = 100%). Power decrease was performed by changing reactor power from 70% to 40%.

The experimental COP task represented performance under normal operating conditions. Operators performed the same four maneuvers in the same manner as described for the initial COP task, with the exception that each team reversed the order in which they used the two types of procedures to perform the COP tasks.

The procedures that the operators used in the study were almost identical to the operators' home plant procedures. Because, operators are typically very well practiced at performing routine operating procedures, they may be so practiced at this task that they perform it almost automatically. Thus, operating the SPWRF in normal operating conditions was likely to be a relatively easy task for the operators. If so, potential performance differences between paper procedures and COPMA-II may not have been evident under normal operating conditions.

Performance under accident conditions was assessed by requiring operators to perform two accident scenarios, a small break loss of cooling accident (LOCA) and a steam generator tube rupture (SGRT). During accident scenarios, operators often must juggle the performance of several procedures. As a result, operators are most likely to be stressed, to become confused, to commit errors, and to need assistance from support systems such as COPMA-II during accident scenarios. Thus, the effects of COPMA-II are most likely to be most evident during performance of the accident scenarios.

2.5.2 Procedure Type Variable

The procedures used in this study were based on the operators' home plant procedures which were modified only to the degree required to make them appropriate for operation of the SPWRF. Operating procedures were written for plant start-up, shutdown, change of reactor power, and for eight accident scenarios. Two of the procedures for accident scenarios were used during the data collection sessions, while the remaining six were used to distract the operators from focusing on the test procedures.

Identical sets of procedures were created in paper and computerized formats. The paper procedures were labeled with their name and number, and hung in the file cabinet beside the SROs work station. The computerized procedures were entered into the COPMA-II data base.

Operators in each team performed the initial and experimental COP tasks with both paper procedures and COPMA-II. However, they performed one accident scenario with paper procedures, and the remaining scenario with COPMA-II.

2.5.3 Team Variable

Each operator performed in only one team throughout the experiment, and the performance and subjective ratings of members of each team were recorded throughout the study to identify any differences between teams. It was important to detect team differences because team differences clouded the results of an evaluation of the first COPMA system. Nelson, Fordestrommen, Holmstrom, Krogsaeter, Karstad, and Tunold [11] found that operators performed more slowly with COPMA than with paper procedures. They suggested that the relatively slow performance with COPMA was not due to characteristics of COPMA, but rather to factors such as operators in some teams being uncommonly cautious with COPMA, or using the COPMA system inefficiently (e.g., double checking parameters that were being automatically monitored).

2.5.4 Trial Block Variable

The trial block variable was used to detect any differences in performance or subjective ratings that were due to the effects of practice. It was important to check for practice effects because, due to practical constraints, such as operator availability, the training sessions during which the operators learned to control the SPWRF with paper procedures and with COPMA-II were relatively short. Short training and/or practice sessions increase the likelihood that the effects of practice are seen in performance measures or subjective ratings.

2.5.5 Control for Order Effects

The order in which the teams performed the initial and experimental COP tasks with paper procedures and with COPMA-II was counterbalanced across teams to prevent order effects in the data. As can be seen in Table 2.1, four teams performed with paper procedures during the first trial of the initial COP task, while the remaining four teams performed these trial with COPMA-II. The order assignment for each team was reversed for the experimental COP task. Thus, members of teams who performed the initial COP task with paper procedures first, performed with COPMA-II first during the experimental COP task trials.

Order effects were also possible during collection of the accident scenario data. Hence, counterbalancing was used to assign team members to the order in which they would perform the two accident scenarios, and to the order in which they would use paper procedures and COPMA-II. Half of the teams performed the LOCA task first, while the other teams performed the SGTR task first. As can be seen in Table 2.2, members in half of the teams that performed the LOCA task first used COPMA-II, while the remaining team members used paper procedures to perform the LOCA task. The same counter-balancing scheme was used for teams that were assigned to perform the SGTR task first.

2.6 Dependent Variables

Two types of dependent variables were measured in the study, performance measures and subjective ratings.

Table 2.1 Type of Procedure Used First by Teams for the Change of Power Tasks

Team Number	Type of Procedure Used First	
	Criterion Change of Power Task	Experimental Change of Power Task
Team 1	COPMA-II	Paper
Team 2	Paper	COPMA-II
Team 3	COPMA-II	Paper
Team 4	Paper	COPMA-II
Team 5	COPMA-II	Paper
Team 6	Paper	COPMA-II
Team 7	COPMA-II	Paper
Team 8	Paper	COPMA-II

Table 2.2 Type of Accident Scenario and Procedure Type Used First by Teams

Team Number	Type of Procedure Used First	
	Loss of Coolant Accident	Steam Generator Tube Rupture Accident
Team 1	COPMA-II	
Team 2	Paper	
Team 3		COPMA-II
Team 4		Paper
Team 5	COPMA-II	
Team 6	Paper	
Team 7		COPMA-II
Team 8		Paper

2.6.1 Response Time Measures

Two aspects of performance in an NPP are particularly important if operators are to respond to accidents or other unexpected events successfully. First, operators must respond quickly. Even more importantly, operators must respond accurately. Hence, time to initiate a procedure (RTI) and time to complete a procedure (RTC) were measured in this study.

For the COP tasks, RTI was defined as the number of seconds elapsed between the SRO's instruction to change reactor power and the time at which the RO pressed the F1 key on the computer keyboard to begin this procedure. RTC was defined as the seconds elapsed between the time at which the RO pressed the F1 key and the time at which reactor power became steady for one minute within the range of 5% above or 5% below the prespecified reactor power. A complete list of timing measures recorded during the COP tasks is provided in Appendix A.

For the LOCA scenario, RTI was defined as the seconds elapsed between the time the researchers initiated the LOCA scenario and the time at which the RO first attempted to trip the reactor. RTC was defined as the seconds elapsed between the time at which the RO first attempted to trip the reactor and the time at which he placed the pressurizer heaters in the "off" position. A complete list of timing measures recorded during the LOCA scenarios can be found in Appendix B.

RTC during the SGTR scenario was defined as the seconds elapsed between the initiation of the SGTR scenario and the time at which the RO first placed the steam dump valve in manual. RTC was defined as the seconds elapsed between the time at which the RO first placed the steam dump valve in manual and the time at which the RO placed the pressurizer heater control in the "off" position. Appendix C Contains a complete list of the timing measures recorded during the SGTR scenario.

2.6.2 Error Measures

The definition of errors for complex tasks is often difficult because there is often more than one way to perform the task correctly. In addition, errors are often additive. For example, if an operator initiates the wrong procedure, that error will result in other incorrect actions (e.g., checking the wrong system parameters, instituting incorrect control responses).

For purposes of this study, errors were defined as any deviation from the optimal sequence of actions within a procedure. To detect such deviations, a prototype of the optimal response sequence was created for each procedure. Thus, the prototype depicted the type and sequence of control actions that allowed operators to complete a procedure in the fewest possible steps. A flow diagram of each operator's control actions was created for each performance trial. The flow diagram was then compared to the appropriate prototype, and the number of control actions made by the operator that deviated from the prototype was recorded as a measure of errors.

2.6.3 Subjective Measures

The NASA-TLX Subjective Workload Scale was used to record operators' subjective ratings of the workload imposed by various conditions. The NASA-

TLX Subjective Workload Scale (see Appendix D), was designed to assess the workload of tasks characterized by cognitive or mental demands [12]. It recognizes the multidimensional nature of mental workload in that it contains rating scales for the magnitude of six workload dimensions: (1) mental demand; (2) physical demand; (3) temporal demand; (4) self-ratings of performance; (5) effort; and (6) frustration. Each of the scales uses a bipolar semantic differential scale format. Thus, operators make ratings on a 100 point scale, with 0 signifying the lowest and 100 signifying the highest workload.

It was important to measure operators' subjective ratings for at least two reasons. First, the advanced aiding systems provided by COPMA-II should reduce mental workload. If COPMA-II organizes information in a more useful fashion, and allows operators to scan procedures more rapidly, operators performance should be superior for COPMA-II trials. The relatively superior performance with COPMA-II is likely to be reflected by lower ratings of mental demand, temporal demand, effort, and frustration for the COPMA-II trials. The operators should also rate their own performance higher for COPMA-II than for paper procedure trials.

Subjective ratings are also important because they provide information about the acceptability and ease-of-use of COPMA-II. Operator acceptance of COPMA-II is essential if COPMA-II is to be used in actual control rooms. If operators dislike new systems, and refuse to use them, even the most advanced technology cannot aid operator performance. Poor operator acceptance has been cited as one of the problems encountered in the use of other advanced systems such as safety parameter display systems [13]

2.7 Research Procedures

Operators were trained and tested at the SPWRF over the course of two days. Upon arrival at the SPWRF, the operators were given a brief description of the study and the data collection schedule. They were then given a brief tour of the SPWRF and the researchers' work station. Before beginning their initial training session, each operator signed a consent form.

2.7.1 SPWRF Training

All training was conducted by the Manager or the Assistant Manager of Training for the SPWRF facility. A part-task training strategy was used in which the operators used paper procedures during the initial training session. Because the SPWRF procedures were based on the operators' home plant procedures, the operators were quite familiar with them, even at the beginning of training. Training operators initially with the familiar paper procedures allowed them to focus most of their attention on learning to operate the SPWRF. Had some operators been trained initially with COPMA-II, those operators would have been forced to divide their attention between learning to operate the SPWRF and learning to operate a new procedure system (COPMA-II).

Training consisted of four one-hour sessions. The first training session employed a lecture/demonstration format in which the operators were familiarized with the thermodynamics and function of the SPWRF, and with the SPWRF control room interface. During the next hour of training, the operators practiced operating the SPWRF with paper procedures. During all practice sessions, the operators practiced by performing a start-up of the SPWRF

(including steaming and drawing the bubble), several COP maneuvers, and a plant shut-down. The third hour of training was a lecture/demonstration in which the operators were familiarized with the structure of COPMA-II, and with the procedures by which they were to interact with COPMA-II. During the fourth training session, the operators practiced operating the SPWRF with COPMA-II.

2.7.2 Research Setting for Training and Data Collection

During the training and experimental sessions, only the control panels of the SPWRF were visible. The remaining components of the SPWRF were enclosed behind a black curtain to ensure that operators did not receive visual cues about system operation that are not available to them at their home plant.

The RO was stationed in the center of the SPWRF console during all training, practice and data collection session. The SRO sat to his right at a small desk that functioned as the SRO work station. The paper procedures were placed in a folder and hung in a file cabinet that was located to the right of the SRO work station. This practice was identical to that used at the operators' home plant. During the COPMA-II trials, the computer on which the COPMA-II system was displayed was placed on the desk at the SRO's work station.

Two video cameras were used to record the operators' performances. One camera was positioned to provide a clear picture of the RO's control movements. The second camera recorded the interaction between the SRO and RO, and the state of the SPWRF annunciators.

The research station was located behind the annunciator panels at the left of the SPWRF console, allowing researchers to function without being seen by the operators. The research station consisted of two VCRs and monitors from which researchers observed operators' actions and recorded the sequence of control movements requested by the SRO and performed by the RO during each task trial.

The response time measures were recorded by two researchers in the research station who recorded the exact times at which operators performed specific actions by monitoring a time stamp located at the lower left corner of the video monitor through which the operators' performance was observed.

To assess the inter-rater reliability of the researchers ratings, the correlation between the ratings of the two research assistants was calculated for response initiation, response completion, and error measures. For cases in which the two raters recorded different timing or error estimates, a mean of the two ratings was used in the data analysis. The inter-rater reliability coefficient for records of response initiation times was $r = .92$, and $r = .94$ for measures of response completion time. The maximum time difference between the two raters for either of the timing measures was 600 ms. The inter-rater reliability coefficient for the accuracy scores was $r = .93$, and the maximum difference between raters for any task trial was one error. These findings indicate that the timing and error data that were collected were reliable and relatively consistent across raters.

2.7.3 Performance of the Change of Power and Accident Scenario Tasks

The initial COP task was performed immediately after the last training session. The operators then performed the experimental COP task that simulated normal operating conditions. The operators in every team performed the accident scenarios last, and performed each accident scenario only one time.

Operators completed the NASA-TLX Subjective Workload Index Scale after they had completed each COP task or accident scenario. Thus, each operator completed the workload scales a total of 10 times. Both the RO and SRO in every team recorded their subjective ratings. They were instructed to write any free-form comments they wanted to make on the back of the workload questionnaires.

2.8 Data Sets and Study Designs

2.8.1 Performance Measures

The performance measures collected for the COP tasks and the accident scenario were identical: (1) RTI; (2) RTC; and (3) number of errors committed. However, the data sets that were collected during performance of the COP tasks and the accident scenarios were different in two main ways. First, the task type variable represented different concepts for the COP tasks and for the accident scenario data sets. For the COP tasks, task type signified the difference between the initial and the experimental COP task. However, accident scenario tasks varied between the LOCA and the SGTR accident scenarios.

The meaning of "trial block" was also different for the COP task and the accident scenario data sets. For the two COP tasks, trial block distinguished between the first set of two COP maneuvers performed (one increasing power, one decreasing power), and the second set. For accident scenarios, the trial block variable referred to whether the accident scenario in question was performed as the first or the last accident scenario.

The data sets for the COP tasks and the accident scenarios were submitted to separate but identical analyses in which the effects of four independent variables were assessed: (1) procedure type; (2) task type; (3) team; and (4) trial block.

2.8.2 Subjective Measures

The subjective measures collected during performance of the COP tasks and the accident scenarios consisted of responses to the six workload scales contained in the NASA-TLX Subjective Workload Scale (see Appendix D). The ratings for each of the six workload scales were submitted to separate but identical analyses in which the effects of five independent variables were analyzed. For the subjective rating data, the effect of operator type (RO vs. SRO) was evaluated as well as the effects of procedure type, task type, team, and trial block.

2.9 Data Analysis

Multivariate Analyses of Variance (MANOVAs) were used to analyze the data initially. MANOVAs are statistical tests that are used to evaluate the effect of one (or more) independent variable on more than one dependent variable. The purpose of the MANOVA is to determine if the effects of any independent variable are significantly different across the multiple dependent measures [14].

When multiple dependent variables are used in a study, the researcher can use a separate evaluation to assess the effects of the independent variables on each dependent measures. However, performing multiple tests makes it more probable that a significant effect will be found due to chance than would be the case if only a single analysis was performed. The increase in the probably of finding a significant effect due to chance is referred to by statisticians as "inflating alpha" [15].

A MANOVA analyzes the data in such as way that alpha is not inflated across the multiple tests, and provides a test of the hypothesis that, for two or more dependent measures, the effects of one or more independent variable(s) is significantly different. The outcome of the MANOVA analysis is a statistic called Wilk's Lambda [14]. A significant Wilk's Lambda score reveals that the effect of one or more independent variables is different across some of the dependent variables. For example, in this study, one significant Wilk's Lambda score indicated that the effects of procedure type was different for RTI and errors than it was for RTC. A nonsignificant Wilk's Lambda score indicates that the effect of the independent variables is the same for all of the dependent measures.

If the Wilkes' Lambda score is significant, the usual practice is to analyze the effects of the independent variables by performing a separate but identical univariate Analyses of Variance (ANOVA) for each of the dependent variables. The "inflated alpha" that will result from the multiple analyses is no longer a concern, because the determination of which independent variables exert a significant effect on which dependent variable has already been determined in the initial MANOVA analysis that controlled the size of alpha across the multiple tests.

The study data were submitted to one of four Multivariate Analyses of Variance (MANOVA) tests. One MANOVA was used to evaluate the effects of procedure type (paper procedures vs. COPMA-II), task type (initial COP, experimental COP, or accident scenario tasks), and team (Team 1 - Team 8) on COP task performance measures. A second MANOVA was used to test the effects of the same factors on the accident scenario data. The third and fourth MANOVAs were used to evaluate the effects of the above mentioned factors, plus the effects of operator type (RO vs. SRO) on subjective workload measures for the COP tasks and the accident scenario tasks, respectively.

3 RESULTS

3.1 Summary of Results

The most important finding of the study was that the operators committed only about half as many errors with COPMA-II as they committed with paper procedures in the accident scenarios (see Figure 3.1). However, accident scenario RTI was fastest for paper procedures, while procedure type did not affect the accident scenario RTC measure. There were no significant effects of procedure type on RTI, RTC, or errors for either the initial or the experimental COP task.

None of the independent variables significantly affected ratings of the six workload dimensions for the accident scenarios. However, the SROs rated temporal demand for the initial COP task higher than did the ROs. For the experimental COP task, the operators rated their own performance higher for COPMA-II than for paper procedures.

The study results are discussed in more detail below.

3.2 Performance Measures

The results of the MANOVA indicated that there were significant differences in the effect of the independent variables on the RTI, RTC, and error performance measures, (Wilks Lambda (7,28) = 4.95, $p < .05$). As a result, separate but identical univariate ANOVAs were performed for each performance measure.

3.2.1 Effects of Procedure Type

The procedure type variable did not exert a significant effect on RTI, RTC, or error for either the initial or the experimental COP task. However, accident scenario errors and RTI were moderated by procedure type. The error difference greatly favored COPMA-II, ($F(1,14) = 16.32$, $p < .01$). The number of accident scenario errors committed with paper procedures was almost twice the number of errors committed with COPMA-II (see Figure 3.1). However, a significant interaction between procedure type and task type, ($F(1,14) = 13.31$, $p < .01$), revealed that the error advantage for COPMA-II was significant only for the LOCA accident scenario data. There was no significant difference in the number of errors committed with the two types of procedures during the SGTR accident scenario.

RTI for the accident scenarios was significantly faster for paper procedures ($m = 208.50$) than for COPMA-II ($m = 297.37$), ($F(1, 14) = 7.47$, $p < .05$). The RTI speed advantage for paper procedures was significant for both of the accident scenarios.

Procedure type did not significantly affect accident scenario RTC overall. However, a significant interaction between procedure type and task type revealed that the effect of procedure type on RTC was different for each of the two accident scenarios, ($F(1, 14) = 7.21$, $p < .05$). The diverging effect of procedure type on RTC for the two accident scenarios is illustrated in Figure 3.2.

For the LOCA scenario, RTC was faster for paper procedures than for COPMA-II, while RTC was fastest for COPMA-II during the SGTR accident scenario.

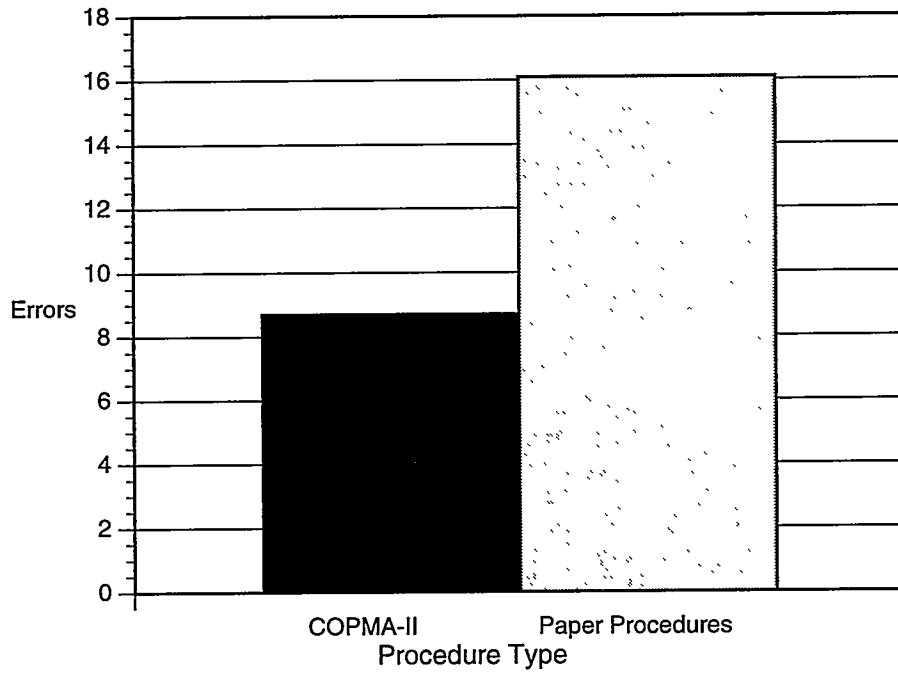


Figure 3.1 Accident Scenario Errors by Procedure Type

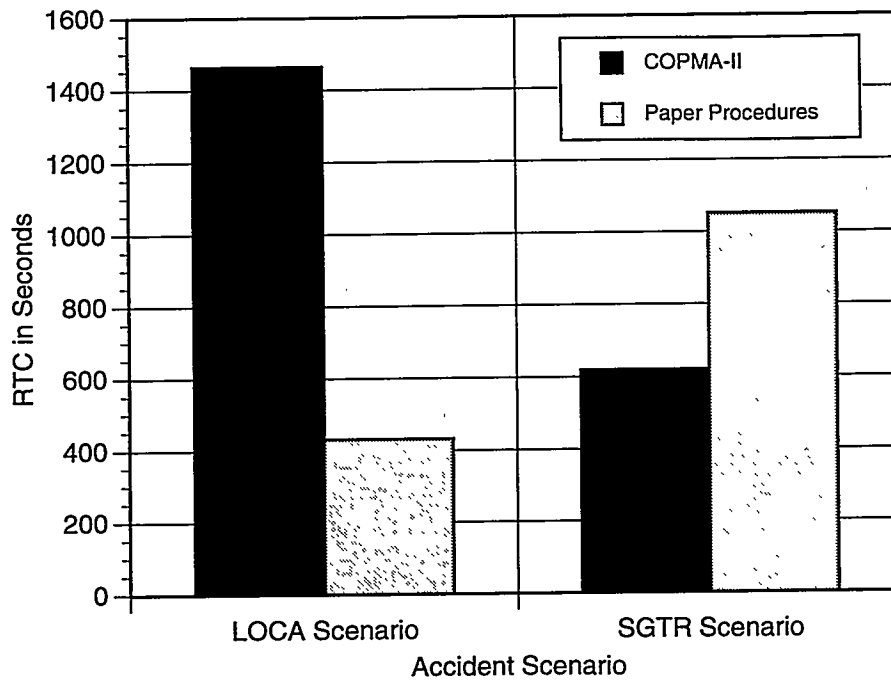


Figure 3.2 RTC for Accident Scenario by Procedure Type

3.2.2 Effects of Trial Block

There were no significant effects of trial block on any of the three performance measures for either the initial or the experimental COP task, or for the accident scenario data.

3.3 Subjective Workload Ratings

A second MANOVA was performed to determine if any of the independent variables affected ratings of the six workload dimensions differently. Some differences were identified, (Wilks Lambda (7,28) = 1.43, $p > .05$). Thus, identical but separate ANOVAs were performed on the rating data for each of the six workload dimensions.

3.3.1 Temporal Demand

There was a significant effect of operator type on ratings of temporal demand for the initial COP task. For the initial COP task, temporal demand was rated markedly higher by SROs ($m = 30$) than by ROs ($m = 17.50$), ($F(1,14) = 4.16$, $p < .05$). Operator type did not significantly affect ratings of any other workload dimension for the initial COP task. In fact, there were no other significant effects of operator type for any combination of workload dimension and experimental task.

3.3.2 Self-Ratings of Performance

For the experimental COP task the operators' self-ratings of performance were higher for COPMA-II ($m = 70.48$) than for paper procedures ($m = 50.64$). There were no other significant differences between self-ratings of performance for any of the remaining workload dimensions, for either the initial or the experimental COP task, or for either of the two accident scenarios.

3.4 Performance Differences Between Teams

The data analyses indicated that the effect of procedure type sometimes varied across teams. For example, the effect of procedure type on RTI for the initial COP task was significantly different across teams, ($F(7,28) = 8.78$, $p < .01$). As can be seen in Figure 3.3, RTI was fastest for paper procedures for four teams, but for one team, RTI was fastest for COPMA-II. RTI was not different across the two types of procedures for the remaining teams.

The effect of procedure type on RTC for the initial COP task was also different across teams, ($F(7,28) = 7.55$, $p < .01$). As can be seen in Figure 3.4, RTC for two teams was fastest with COPMA-II, while RTC for two other teams was fastest with paper procedures. RTC for the remaining teams was not significantly different for COPMA-II and paper procedures.

The relationship between procedure type and accident scenario RTC was significantly different across teams, ($F(1,14) = 18.93$, $p < .001$). As can be seen in Figure 3.5, five teams completed accident scenario responses fastest with paper procedures, while one team completed accident scenario responses fastest with COPMA-II. For the remaining two teams, RTC was not significantly different across the two procedure types.

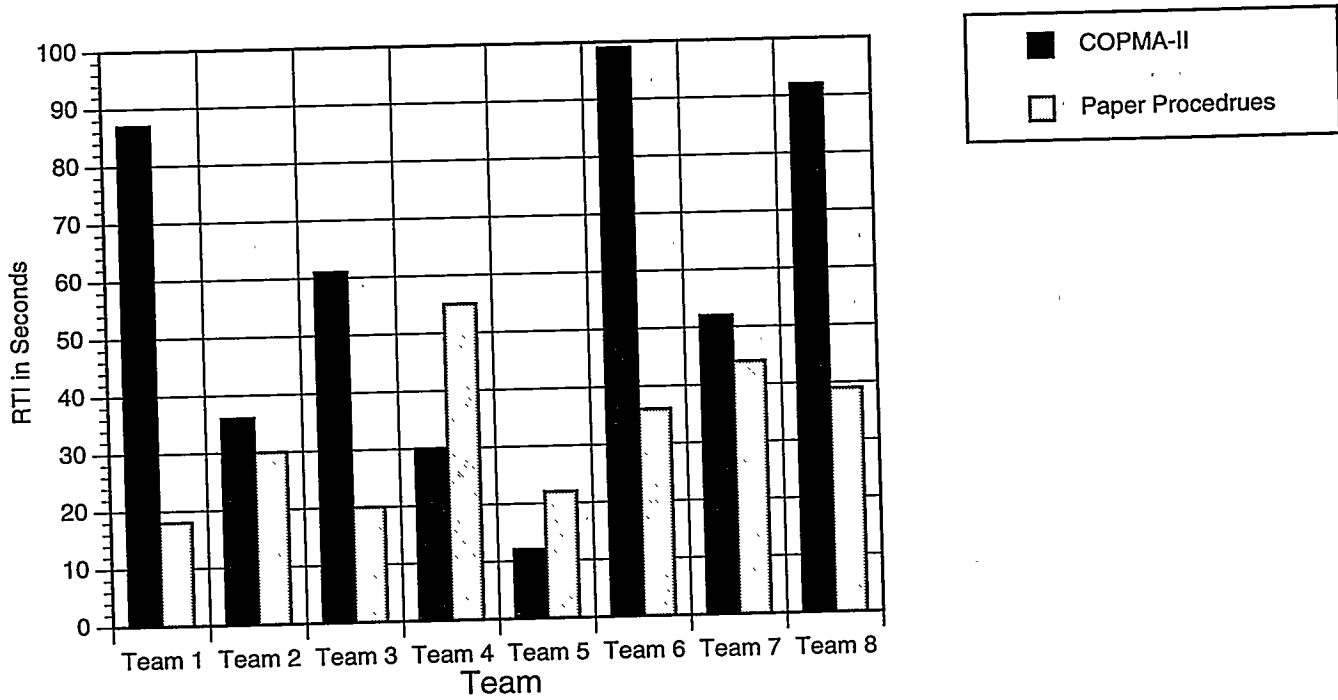


Figure 3.3 Initial Change of Power Task RTI for Procedure Type by Team

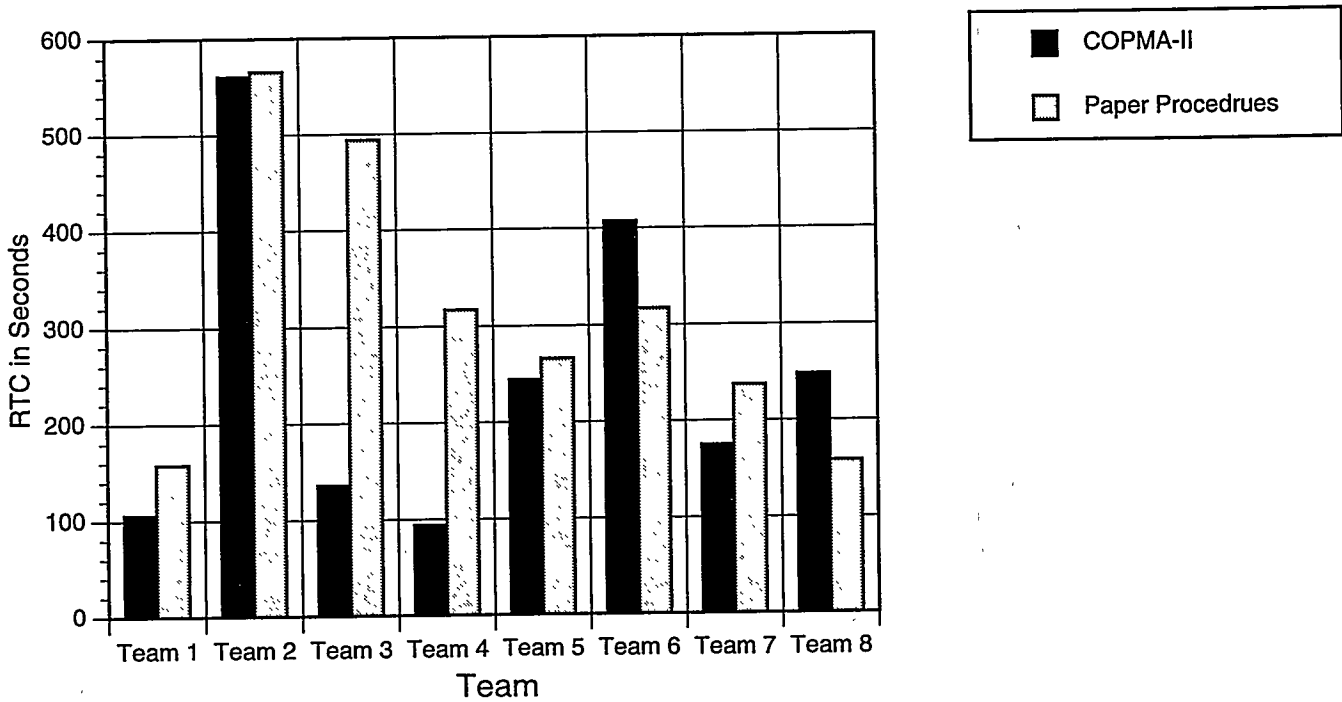


Figure 3.4 Initial Change of Power Task RTC for Procedure Type by Team

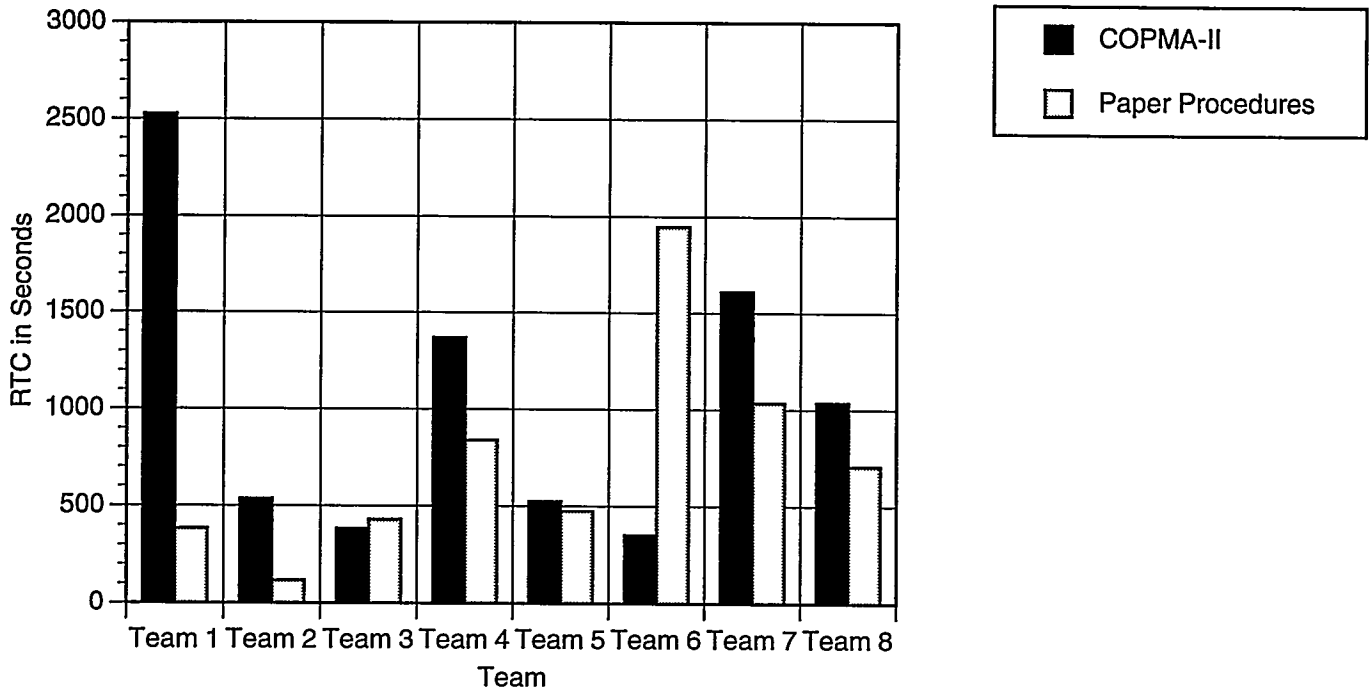


Figure 3.5 Accident Scenario RTC for Procedure Type by Team

The effect of procedure type on the number of accident scenario errors was significantly different across teams, ($F(7,14) = 8.33, p < .01$). As can be seen in Figure 3.6, three teams committed the fewest errors with COPMA-II, while the number of errors committed by members of the remaining teams was not significantly different for paper procedures and COPMA-II.

3.5 Differences in Subjective Ratings Between Teams

3.5.1 Effect of Procedure Type on Individual Teams

In the initial COP task, the ratings of mental demand for the two procedure types were significantly different across teams, ($F(7, 14) = 14.29, p < .05$). As can be seen in Figure 3.7, members of Team 8 rated mental demand highest for paper procedures, while members of Teams 2 and 7 rated mental demand highest for COPMA-II. The ratings of the mental demand of the initial COP task were not significantly different across the two procedure types for the remaining teams.

The effect of procedure type on ratings of frustration for the initial COP task also varied across teams, ($F(7,14) = 8.78, p < .01$). Members of Teams 2 and 8 rated frustration highest for paper procedures, while members of Teams 1 and 7 rated their frustration highest for COPMA-II (See Figure 3.8). Members of the remaining teams did not rate frustration differently across the two types of procedures.

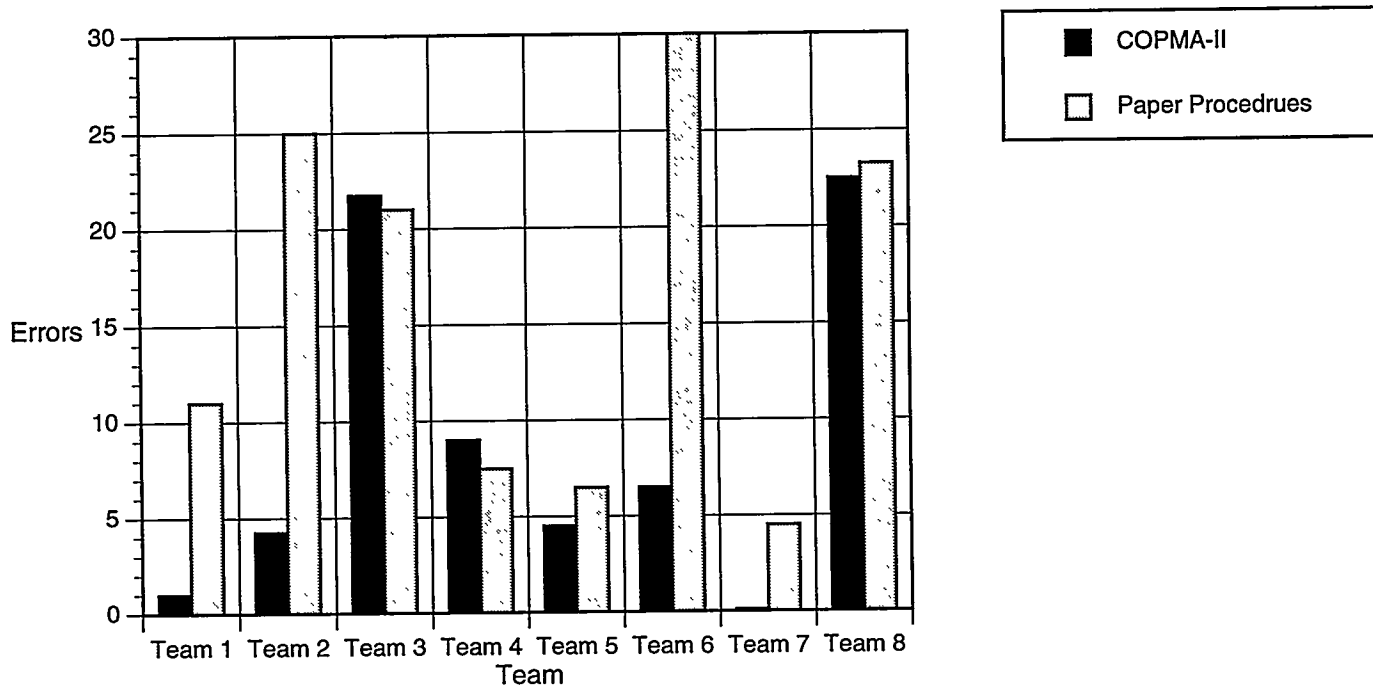


Figure 3.6 Accident Scenario Errors for Procedure Type by Team

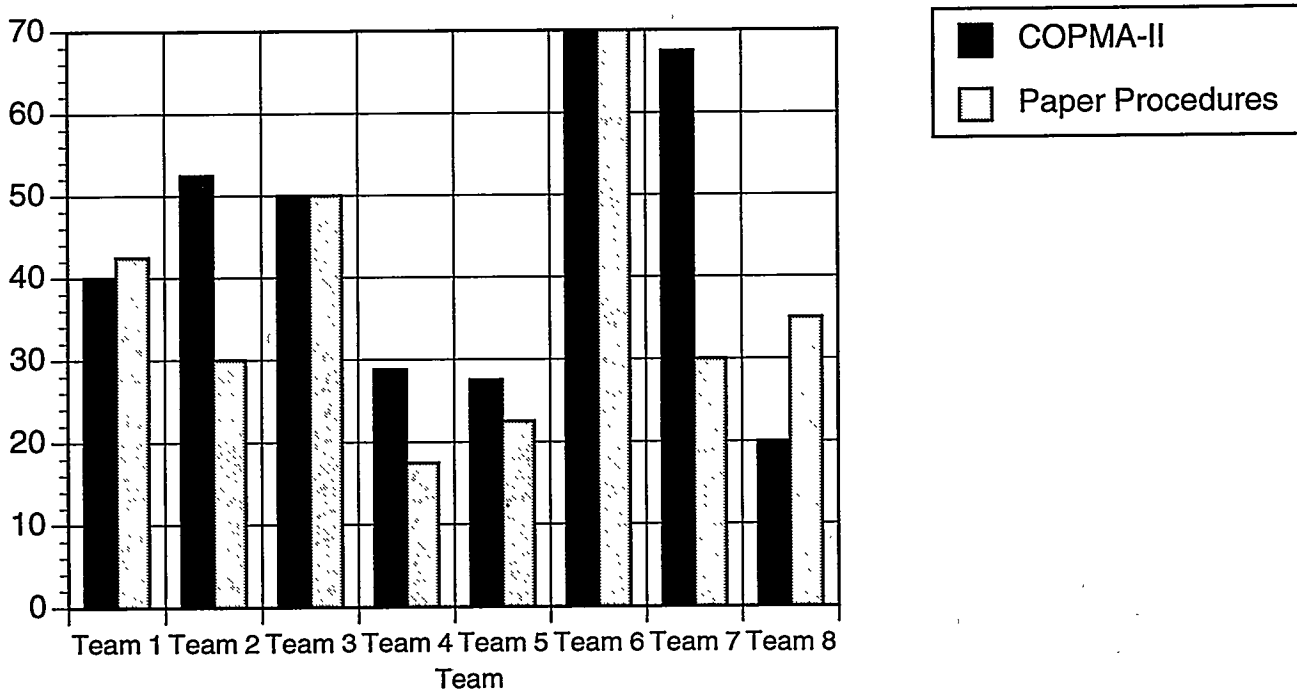


Figure 3.7 NASA-TLX Ratings of Mental Demand for Initial Change of Power Task for Procedure Type by Team

3.5.2 Other Team Differences in Subjective Ratings of Workload

Frustration experienced in the experimental COP task was rated differently by the members of various teams, ($F(7,14) = 15.83, p < .01$). This finding is illustrated in Figure 3.9. In terms of COP task ratings of frustration, the scores of team members fell into four distinct groups, with each group representing a distinct level of frustration. The first group is represented by only Team 1, the members of which rated frustration highest overall. The remaining three groups, in descending order of the magnitude of their frustration ratings, were Teams 2, 6, and 7; Teams 4, 5, and 8; and Team 3.

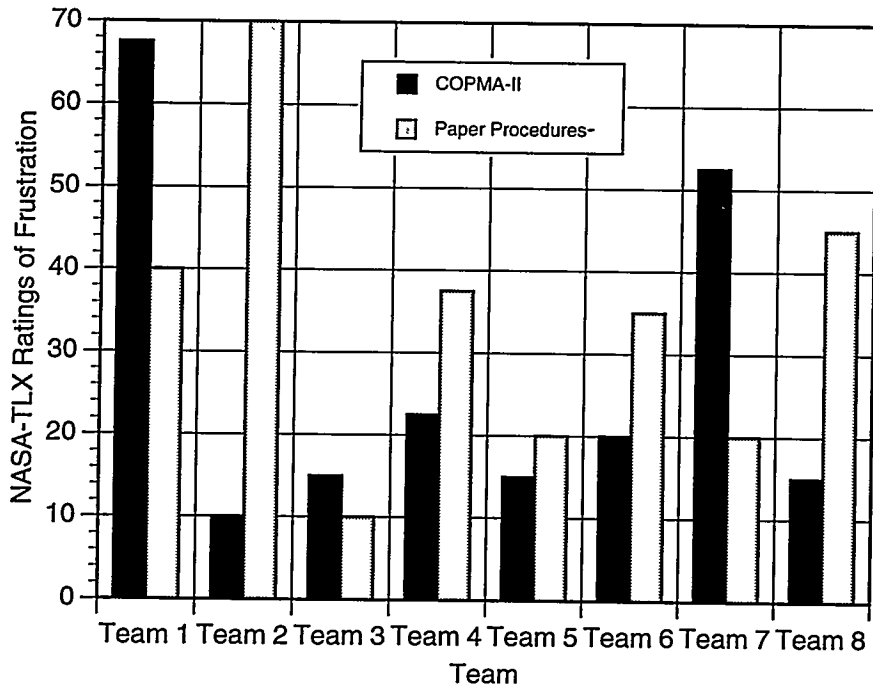


Figure 3.8 Initial Change of Power Task NASA-TLX Ratings of Frustration for Procedure by Team

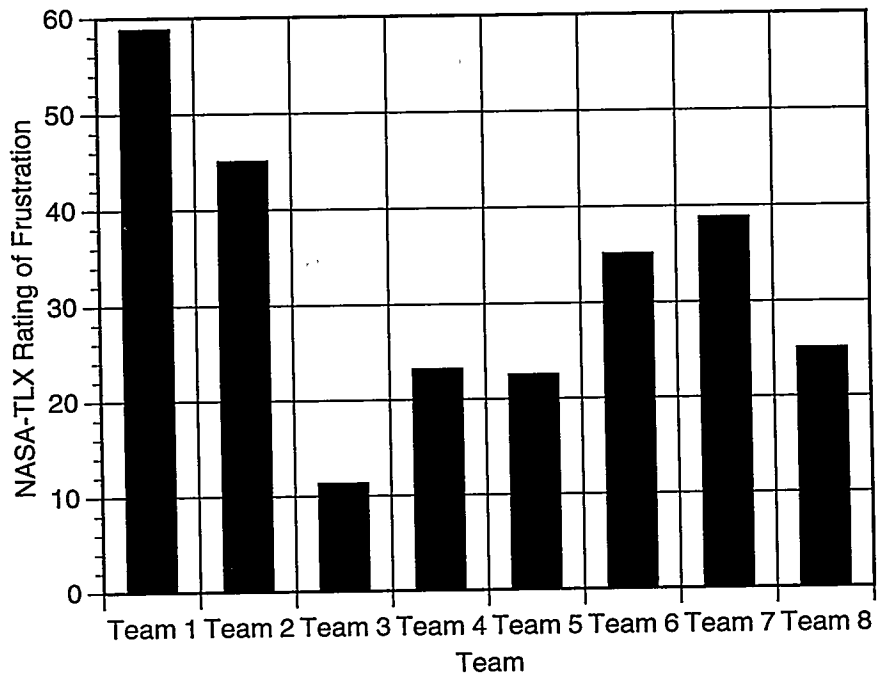


Figure 3.9 Experimental Change of Power Task NASA-TLX Ratings of Frustration by Team

4 DISCUSSION OF STUDY RESULTS

The following is a discussion of: (1) the study findings as related to the study purpose and initial hypotheses as well as potential causes of unexpected results; and (2) the results of the study in terms of lessons learned about the research methods that should be used to validate proposed system changes such as computerized operating procedures.

4.1 Study Objectives and Hypotheses

The objective of this study was to assess the relative effects of paper procedures and a computerized procedure system, COPMA-II, on the speed and accuracy of the control maneuvers of licensed reactor operators. The second study objective was to assess operator acceptance of the COPMA-II system by comparing operators' subjective ratings of mental workload across performance with paper procedures and COPMA-II. The study was designed in such a way that performance and subjective ratings could be compared across paper procedures and COPMA-II, across normal operating conditions and accident scenarios, across time, and across teams of operators.

It was hypothesized that performance with COPMA-II would be more rapid and contain less mistakes than performance with paper procedures. However, this result was predicted only for performance of the accident scenarios. Due to the relative ease of the Change of Power (COP) Task that simulated normal operating conditions, a difference in performance measures due to procedure type was not predicted for the COP task. No significant differences in either performance or subjective workload ratings were predicted across time, or across teams of operators.

4.2 Accident Scenario Data

4.2.1 Error Data

As predicted, there was no significant difference between time to initiate a response (RTI), time to complete a response (RTC), or errors for paper procedures and COPMA-II on the COP task. The hypothesis that, for the accident scenario data, there would be a significant difference in the number of errors committed with paper procedures and COPMA-II was also confirmed. The most important finding of the study was that there were almost twice as many accident scenario errors with paper procedures than with COPMA-II. The COPMA-II error advantage is especially impressive because it was obtained when operators had only a relatively small amount of practice with COPMA-II.

Our data do not provide a definitive explanation of the error advantage obtained for COPMA-II. However, comments made by several operators provide some possible explanations. Several operators commented that COPMA-II did not allow them to "skip ahead" and preview future procedure steps, as they could do with paper procedures. The paper procedures used in this study were quite similar to the operators' home plant procedures, and the operators seemed to become comfortable with them almost immediately. Thus, the operators reported that they were able to move back and forth in the paper procedures relatively easily. Because any deviation from the predefined optimal sequence of procedure steps was counted as an error, temporarily leaving the predefined procedure path to look ahead constituted an action that was counted as an error. If operators did look ahead when using paper procedures, but were

forced to be more methodical when using COPMA-II, the error rate would be highest for paper procedures.

The COPMA-II error advantage that was obtained in this study does not necessarily mean that operators will make the fewest errors with COPMA-II in operational settings. It may be that experienced operators perform rapidly enough to preview and prepare for future procedural steps without degrading the quality of their performance. In fact, this practice may be an important component of the skilled performance of experienced operators.

In an operational setting, operators would probably learn to maneuver in COPMA-II with as much ease as they now maneuver through paper procedures. This development could possibly erase the error advantage for COPMA-II. The current configuration of COPMA-II does include advanced aiding features including a mechanism that allows operators to deviate from a preestablished procedure path. However, the operators who performed in this study were not trained to use these features because of the rather restricted practice and training time allotted for practice on COPMA-II, and because differences in the layout and content of the COPMA-II interface across teams of operators would make interpretation of the study results difficult.

While the overall number of accident scenario errors was much lower for COPMA-II than for paper procedures, when the error rates were examined separately for each individual accident scenario, it was clear that the error advantage for COPMA-II existed only for the LOCA scenario data. The reason that COPMA-II provided an error advantage for the LOCA, but not for the SGTR scenario, is not clear. The effects of practice cannot account for this finding because half of the teams performed the relatively accurate LOCA scenario first. The LOCA error advantage for COPMA-II could indicate that the LOCA scenario was more difficult to perform than was the SGTR scenario, and that the greater demands of the LOCA task made this task more sensitive to the effects of procedure type. However, the procedures that the operators used to perform the LOCA and the SGTR scenarios were quite similar, and none of the operators commented that either of the two accident scenarios was more challenging than the other. What the disparate effect of the two accident scenarios on errors does clearly illustrate is that task type is an important factor that should be examined in studies that assess the validity of computerized procedures.

4.2.2 Response Time Data

While COPMA-II reduced the number of accident scenario errors, accident scenario RTI was slowest for COPMA-II. Thus, the accident scenario error and RTI data describe a speed-accuracy tradeoff in which COPMA-II reduced the number of errors, but lengthened the time required to initiate responses. However, the speed-accuracy trade-off engendered by COPMA-II probably does not indicate that COPMA-II would seriously degrade operator performance in operational conditions. Errors are usually a more critical factor in recovery from accidents or transients than is the speed with which emergency operating procedures are initiated. Thus, the finding that there were the least errors committed with COPMA-II is probably more meaningful than the finding that RTI was slower for COPMA-II than for paper procedures.

The comments of some operators suggested that the slowed RTI for accident scenarios was partially generated by specific characteristics of the COPMA-II interface. Several operators noted that it was difficult to initiate a

procedure while looking only at the relatively small part of a procedure that appeared in the Current Instruction Pane on the COPMA-II Interface. Thus, the operators found it necessary to stop to expand the size of the Current Instruction Pane before beginning execution of the procedure. If SROs stopped to make this adjustment on the COPMA-II interface before they directed the RO to press the F1 key (thus, concluding the measurement of RTI), the RTI for COPMA-II trials would be delayed relative to that for paper procedures.

Another relevant comment made by operators was that the response time of COPMA-II was relatively slow and, at times, it was necessary to wait for the requested procedures to appear on the COPMA-II Instruction Pane. If SROs had to wait for instructions when they used COPMA-II, but were able to begin reading paper procedures immediately, the RTI for COPMA-II trials would be longer than the RTI for paper procedures. However, the finding that there was no difference between RTC for paper procedures and for COPMA-II argues against this explanation of the slow RTI for COPMA-II because a slow system response time should increase COPMA-II RTC as well as RTI.

Several operators remarked that the boxes on which they were to click the mouse when using COPMA-II were so small that it was often necessary to click several times before successfully opening a procedure. If SROs had to click several times before successfully opening a COPMA-II procedure, RTI for COPMA-II trials would be increased because the SRO could not instruct the RO to press the F1 key until the procedure was successfully opened. However, this explanation of the delayed RTI for COPMA-II is also challenged by the finding that accident scenario RTC was not different for COPMA-II and paper procedures. The SROs were required to click on the mouse boxes throughout the accident scenarios. Thus, the small size of the mouse box would increase COPMA-II RTC as well as RTI.

4.3 Subjective Workload Data

The second purpose of this study was to assess operator acceptance of COPMA-II by comparing operators' ratings of subjective workload between COPMA-II and paper procedure trials. Due to the relative ease of the COP task that simulated normal operating conditions, it was not predicted that ratings of mental workload would be different for paper procedures and COPMA-II. The majority of the subjective workload data support this hypothesis.

The only significant difference in subjective workload ratings for the initial COP task was that temporal demand was rated higher by SROs than by ROs. This finding provides evidence of the validity of the NASA-TLX workload measures used in this study. Because SROs typically carry more responsibility than do ROs, they are likely to experience the most time pressure. Thus, the magnitude of the SROs' ratings of temporal load should be greater than the magnitude of the ratings of the ROs. It is also important to note that, while SROs experienced the most time pressure during the initial COP task, the time pressure they experienced was not moderated by procedure type; the temporal demand ratings of the SROs was no different for paper procedures and for COPMA-II for the initial COP task.

For the experimental COP task, the only significant effect of procedure type on subjective workload ratings was that operators were most confident of their performance when they used COPMA-II. This result is consistent with operators' comments that COPMA-II structured their responses and made it less

likely that a required action would be missed, or that procedure steps would be performed in an incorrect order. It is interesting to note that the operators self-ratings of performance did not accurately predict the effects of procedure type on the number of errors committed during the experimental COP task. During the experimental COP task, there was no significant difference between the number of errors committed with paper procedures and with COPMA-II.

The finding that, for the accident scenarios, there were no significant differences between ratings of mental workload for the two procedure types for any of the six dimensions of mental workload indicates that once the operators had completed the COP tasks, they did not experience different amounts of mental workload for each of the two procedure types during the accident scenarios. The majority of verbal comments made by operators after performing the two accident scenarios were favorable to COPMA-II, and several operators were noticeably enthusiastic about using COPMA-II in their actual control rooms.

4.4 Research Methods and Lessons Learned

The findings of this study provide some important information about the research methods that should be used to validate devices such as computerized operating procedures. The following discussion will address four main methodological issues. The first issue concerns the types of task that should be used to validate devices such as computerized procedures. The second issue is the selection of performance measures. The third issue concerns the amount of time that operators should be allowed to practice with new devices before validation data are collected. Finally, the effect of differences in skill level between operators will be considered.

4.4.1 Effects of Task Type

The results of the study indicate that some tasks may be sensitive to differences in the speed and error rate of operator performance while other tasks are not. For example, the effect of procedure type on performance data for the LOCA and SGTR accident scenarios was quite different. The cause of the marked differences in the effect of procedure type on accident scenario performance measures is difficult to ascertain and cannot be conclusively defined by the data provided by the current study. However, this finding does clearly indicate that more than one type of task should be used to validate systems such as computerized procedures. The use of several tasks in validation studies would not only be likely to increase the validity of study results, but would also help researchers identify the key features of various tasks that render computerized procedures helpful in some situations, but neutral or even harmful to performance in others.

The finding that there were no significant effects of procedure type on the performance of the COP task, but that procedure type did moderate performance of the accident scenarios, indicates that task difficulty is one task characteristic that is important for exposing the effects of computerized procedures on operator performance. The clear error advantage for COPMA-II was operative only for data collected during the relatively difficult accident scenarios. If operators had performed only the relatively simple COP tasks, the robust error advantage for COPMA-II would not have been discovered. Thus, it is important that operators perform relatively difficult tasks to

demonstrate potential performance differences between performance with paper procedures and computerized systems.

4.4.2 Performance Measures

Many of the operators who performed this study attributed the accident scenario error advantage for COPMA-II to the reduced likelihood that they would skip a procedure step or commit a sequence error when they used COPMA-II. As mentioned previously, previewing procedures may be an effective way of preparing to perform up-coming activities. Thus, the way errors were defined in this study may not have reflected differences in performance that were truly important to the quality of performance in operational settings. The lesson learned from this possibility is that the precise measures used in validation studies should not only be carefully considered, but should also be thoroughly discussed with task experts in terms of their meaningfulness to real-world settings.

4.4.3 Effects of Practice

Another lesson learned in the current study concerns the amount of time operators should be allowed to train and practice with new systems before validation data are collected. Due to the work schedules of the operators who performed the current study, it was possible to employ them for only two days. This constraint dictated that the time allotted for training and practice be limited to one eight-hour day. Due to this time constraint, it was possible to train operators to use COPMA-II in only its most simple default configuration.

It is interesting to consider whether the speed advantage for the paper procedures obtained in this study would diminish as operators became well practiced with using COPMA-II. In the current version of COPMA-II, it is possible to use an expanded Current Instruction Pane as a default setting. However, the operators who performed in this study were not trained to use these COPMA-II features because of the restricted training and practice time, and because differences in the layout and content of the COPMA-II interface across teams of operators could confound the interpretation of the study results. However, if operators had been trained to use these advanced COPMA-II features, the speed advantage for the paper procedures may not have occurred. On the other hand, the error advantage for COPMA-II might also have disappeared with practice because the operators may have learned to preview future procedure steps as easily with COPMA-II as they now do with paper procedures.

4.4.4 Differences in the Skill Level of Operators

In the evaluation conducted by Halden of the initial COPMA system, large differences in the initial skill level of the operators made interpretation of the data difficult [11]. To avoid the confounding effects of individual differences in the current study, criterion data were collected at the end of the training period, while the operators performed the initial COP task. This procedure was based on the assumption that any differences in the skill level of operators would be apparent at the end of training, and that if the operators' performance was not significantly different at the end of training there were no important differences in the skill levels of operators.

Unfortunately, the findings of the current study indicate that the assumption that similarity in the level of operator performance at the end of training indicates that there are no important differences in operator skill levels was questionable. While there were no substantial differences in the performance of teams for the criterion or experimental COP task, the performance of teams varied significantly during performance of the accident scenarios. For example, the effect of procedure type on accident scenario performance was significantly different across the teams. For Teams 1, 2, and 6, the number of accident scenario errors was higher for paper procedures than for COPMA-II. However, there was no significant difference in the number of accident scenario errors for paper procedures and for COPMA-II for the remaining teams. Thus, while the accident scenario error advantage for COPMA-II was robust, it was operative for less than half the teams.

The finding that the number of errors committed with COPMA-II varied significantly between teams indicates that the effect of factors such as procedure type may be different for different teams when operators perform relatively difficult tasks. Another example of important performance differences between teams was found in the RTI data for the accident scenario tasks where half of the teams initiated their responses faster with paper procedures than with COPMA-II. However, there was no difference in RTI between the two types of procedures for the rest of the teams. These findings, in combination with the lack of significant differences between the performance of teams for the COP tasks, suggests that, even when teams are tested for equal performance before data collection (at the end of training), there may be differences in the skill levels of operators that do not become obvious until operators perform challenging or difficult tasks such as accident scenarios.

It appears that testing for differences in skill levels between operators is not an effective way to control for individual differences between operators. Rather, the most effective and reliable method of controlling the effects of individual differences is probably to increase the number of operators who perform in validation studies to a number of operators that is great enough to eliminate the effects of individual differences by randomly assigning operators to various study conditions. Thus, another lesson learned in this study was that the number of operators who perform in validation studies should be increased from the number that is now typically used in current studies to a sample size of at least 20 - 30 operators.

5 REFERENCES

- [1] P.C. Goodman, & C.A. DiPalo, "Human factors information system: A tool to assess error related to human performance in U.S. nuclear power plants." Proceedings of the Human Factors Society 35th Annual Meeting. October, 1991, 662-665.
- [2] G. West, Jr., R.J. Eckenrode, & P.C. Goodman, "Investigation of events involving human performance." Proceedings of the Human Factors Society 35th Annual Meeting. October, 1991, 655-658.
- [3] M. Paradies, L. Unger, & A. Ramey-Smith, "NRC Human Performance Investigation Process (HPIP)." Proceedings of the Human Factors Society 35th Annual Meeting. October, 1991, 650-654.
- [4] C.A. Tolbert, C.J. Moore, & D.R. Wierings, "Emerging issues for procedures in the nuclear industry." Proceedings of the Human Factors Society 35th Annual Meeting. October, 1991, 1238-1242.
- [5] C.S. Isakson, & J.H. Spyridakis, "The comprehensibility of linear versus nonlinear documents. In TC Toolbox Proceedings Third Region 7 Conference of the Society for Technical Communication, October, 1990.
- [6] N.T. Fordestrommen, & K. Haugset, "Integrating surveillance and control without overwhelming the operator." Nuclear Engineering International, September, 1991, 41-43.
- [7] K.W. Olsen, J. Teigen, E. Ness, "MMI design proposals for COPMA-II On-Line," OECD Halden Reactor Project MMSR-1232, November, 1991.
- [8] J. Tiegen, E. Ness, F. Flandelsby, "Computerizing operating procedures with COPMA-II: A disturbance procedure from paper to COPMA-II implementation," OECD Halden Reactor Project MMSR-1232, February, 1993.
- [9] M. Krogsaeter, J.S. Larsen, S.R. Nilsen, & F. Owre, "The Computerized Procedure System COPMA - System description and user interface," OECD Halden Reactor Project HPR-337, June, 1989.
- [10] W.T. Bilger, & J.C. Deardorff, "A model project," The Nuclear Professional. Vol. 5, No. 4, 23-25 (1990).
- [11] W.R. Nelson, N.T. Fordestrommen, C.B.O. Holmstrom, M. Krogsaeter, T. Karstad, & O. Tunold, "Experimental evaluation of the computerised procedure system COPMA," OECD Halden Reactor Project report HWR-277, December, 1990.
- [12] M.A. Vidulich & P.S. Tsang, "Assessing subjective workload assessment: A comparison of SWAT and the NASA-Bipolar methods." Proceedings of the Human Factors Society 29th Annual Meeting. October, 1985, 71-75.
- [13] S.H. Weiss, W.H. Regan, & J.W. Roe, "Experience with operator aids for nuclear power plants in the United States of America," p. 323-329, in Man-machine interface in the nuclear industry. Conference proceedings of an international conference of man-machine interface in the nuclear industry. International Atomic Energy Agency, 1988.

[14] N. Cliff, Analyzing multivariate data, p. 253-254, Harcourt Brace Jovanovich, San Diego, CA., 1987.

[15] L.A. Marascuito & J.R. Levin, Multivariate statistics in the social sciences: A researchers guide. p. 184, Monterey, CA., Brooks/Cole.

APPENDIX A

TIMING MEASURES RECORDED DURING THE CHANGE OF POWER TASKS

Timing Measures Recorded During the Change of Power Tasks:

1. Time at which a researcher began to instruct the senior reactor operator to change reactor power.
2. Time at which the reactor operator pressed the F1 Key on the computer keyboard.
3. Time at which reactor power first reached power between 40kw - 50kw (descending power change) or between 65kw - 75kw (ascending power change).
4. Time at which reactor power remained steady for one minute at a point between 40kw - 50kw (descending power change) or between 65kw - 75kw (ascending power change).
5. Time senior reactor operator lifted phone to report stabilization time.

APPENDIX B

TIMING MEASURES RECORDED DURING THE LOSS OF COOLANT ACCIDENT SCENARIO

Timing Measures Recorded During the Loss of Coolant Accident Scenario

1. Time at which the switch to begin the loss of cooling accident scenario was switched by a member of the research team.
2. Time at which the senior reactor operator began to announce the occurrence of a small break loss of cooling accident on paging system.
3. Time at which the senior reactor operator began to read the EP/001.
4. Time at which the reactor operator first attempted to trip the reactor.
5. Time at which the reactor operator manually tripped the reactor.
6. Time at which the senior reactor operator announced the time at which the reactor was tripped.
7. Time at which the senior reactor operator began to read EP 1.3.
8. Time at which the reactor operator placed the steam dump valve in manual.
9. Time at which the reactor operator placed the pressurizer heater control in the off position.
10. Time at which the T Average setting reached 170 degrees.

APPENDIX C

TIMING MEASURES RECORDED DURING THE STEAM GENERATOR TUBE RUPTURE
ACCIDENT SCENARIO

Timing Measures Recorded During the Steam Generator
Tube Rupture Accident Scenario

1. Time at which the switch to begin the loss of steam generator tube rupture accident scenario was switched by a member of the research team.
2. Time at which the senior reactor operator began to announce the occurrence of a primary on the paging system.
3. Time at which the senior reactor operator began to read the EP/001.
4. Time at which the senior reactor operator lifted the phone to check with the Facility Manager for further actions required.
5. Time at which the reactor operator manually placed the steam dump valve in manual.
6. Time at which the reactor operator placed the pressurizer heater control in the off position.
7. Time at which the system T Average setting reached 170 degrees.

APPENDIX D

NASA-TLX SUBJECTIVE WORKLOAD SCALE:
INSTRUCTIONS READ TO SUBJECTS AND PROTOTYPE SCALE

NASA-TLX Subjective Workload Scale:
Instructions Read To Subjects and Prototype Scale

We are interested in assessing the experience you will have during the different task scenarios. Right now I am going to describe the technique that will be used to examine your experience. In the most general sense, we are examining the "workload" you will be experiencing. Workload is a difficult concept to define precisely, but a simple concept to understand generally. The factors that influence your experience of workload may come from the task itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you felt. The workload contributed by different task elements may change as you get more familiar with the task, perform easier or harder versions of it, or move from one task to another. Physical components of workload are relatively easy to conceptualize and evaluate. However, the mental components or workload may be more difficult to measure.

Since workload is something that is experienced individually by each person, there are not effective "rulers" that can be used to estimate the workload of different activities. One way to find out about workload is to ask people to describe the feelings they experienced. Because workload may be caused by many different factors, you will be asked to evaluate several factors individually rather than lumping them into a single global evaluation of overall workload.

(Experimenter presents operator with the Rating Scale Definition Table.)

This set of six rating scales was developed for you to use in evaluating your experience during different tasks. Please read the descriptions of the scales carefully. If you have any questions about any of the scales in the table, please ask me about it. It is extremely important that they be clear to you. You may keep the description table with you for reference when you are completing the scales.

After performing each task scenario, you will be presented with these six rating scales. You will evaluate the tasks by making a mark on each of the six scales that matches your experience. Each scale has two endpoint descriptors that describe the scale. Please consider your responses carefully, and try to use the scales to distinguish among the different task elements. Consider each scale individually. The ratings you provide will play an important role in the evaluation that is being conducted.

Rating Scale Definitions

1. Mental Demand (Low Versus High)

How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?

2. Physical Demand (Low Versus High)

How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy, demanding, restful or laborious?

3. Temporal Demand (Low Versus High)

How much time pressure did you feel due to the rate or pace at which the task elements occurred? Was the pace slow and leisurely or rapid and frantic?

4. Performance (Good Versus Poor)

How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

5. Effort (Low Versus High)

How hard did you have to work (mentally and physically) to accomplish your level of performance?

6. Frustration Level (Low Versus High)

How discouraged, irritated, stressed or annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

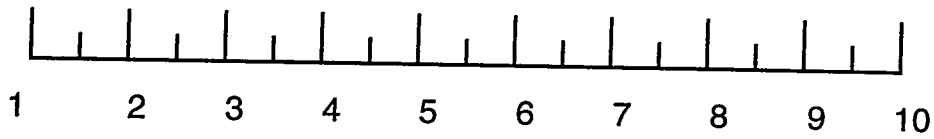
Team #: _____

Scenario #: _____

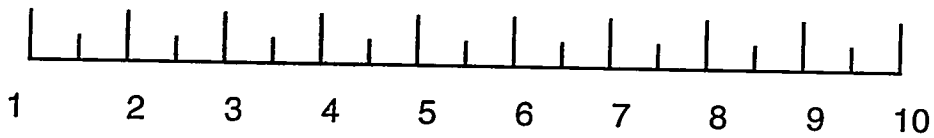
Copma Condition: _____

Scenario Type: _____

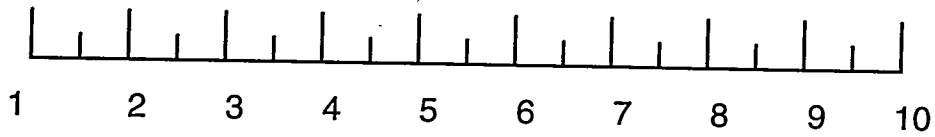
Mental Demand



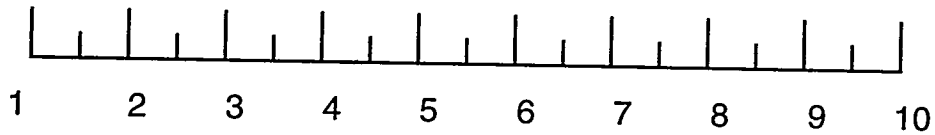
Physical Demand



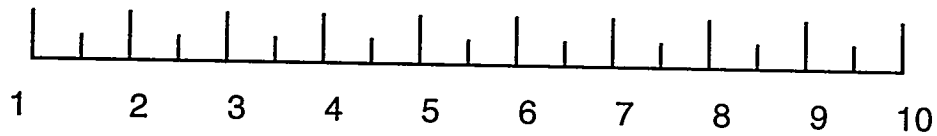
Temporal Demand



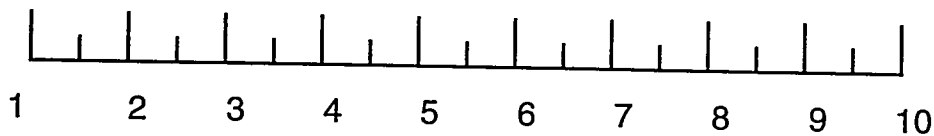
Performance



Effort



Frustration



Comparison Task

On each line below, you will find the names of two aspects of workload. On each line, please circle the concept that you think contributed the most to your overall workload.

- | | | | |
|-----|--------------------|---|-----------------|
| 1. | Effort | - | Performance |
| 2. | Temporal Demand | - | Effort |
| 3. | Performance | - | Frustration |
| 4. | Physical Demand | - | Performance |
| 5. | Temporal Demand | - | Frustration |
| 6. | Physical Demand | - | Frustration |
| 7. | Physical Demand | - | Temporal Demand |
| 8. | Temporal Demand | - | Mental Demand |
| 9. | Frustration | - | Effort |
| 10. | Performance | - | Temporal Demand |
| 11. | Mental Demand | - | Physical Demand |
| 12. | Frustration | - | Mental Demand |
| 13. | Performance Demand | - | Mental |
| 14. | Mental Demand | - | Effort |
| 15. | Effort | - | Physical Demand |

BIBLIOGRAPHIC DATA SHEET

(See instructions on the reverse)

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

NUREG/CR-6398

2. TITLE AND SUBTITLE

Evaluation of the Computerized Procedures Manual II (COPMA-II)

3. DATE REPORT PUBLISHED

MONTH | YEAR

November | 1995

4. FIN OR GRANT NUMBER

11979

5. AUTHOR(S)

S. A. Converse

6. TYPE OF REPORT

7. PERIOD COVERED (Inclusive Dates)

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

North Carolina State University
Raleigh, NC 27695-7801

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

Division of Systems Technology
Office of Nuclear Regulatory Research
U. S. Nuclear Regulatory Commission
Washington, D. C. 20555-0001

10. SUPPLEMENTARY NOTES

J. Persensky, NRC Project Manager

11. ABSTRACT (200 words or less) The purpose of this study was to evaluate the effects of a computerized procedure system, the Computerized Procedure Manual II (COPMA-II), on the performance and mental workload of licensed reactor operators. To evaluate COPMA-II, eight teams of two operators were trained to operate a scaled pressurized water reactor facility (SPWRF) with traditional paper procedures and with COPMA-II. Following training, each team operated the SPWRF under normal operating conditions with both paper procedures and COPMA-II. The teams then performed one of two accident scenarios with paper procedures, but performed the remaining accident scenario with COPMA-II. Performance measures and subjective estimates of mental workload were recorded for each performance trial. The most important finding of the study was that the operators committed only half as many errors during the accident scenarios with COPMA-II as they committed with paper procedures. However, time to initiate a procedure was fastest for paper procedures for accident scenario trials. For performance under normal operating conditions, there was no difference in time to initiate or to complete a procedure, or in the number of errors committed with paper procedures and with COPMA-II. There were no consistent differences in the mental workload ratings operators recorded for trials with paper procedures and COPMA-II.

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

Computerized Procedures
Procedures
Human Factors
Workload
Nuclear Power Plant Procedures

13. AVAILABILITY STATEMENT

Unlimited

14. SECURITY CLASSIFICATION

(This Page)

Unclassified

(This Report)

Unclassified

15. NUMBER OF PAGES

16. PRICE