AL/CF-TR-1994-0040

ARMSTRONG LABORATORY

MICROGRAVITY EFFECTS ON COGNITIVE PERFORMANCE MEASURES: PRACTICE SCHEDULES TO ACQUIRE AND MAINTAIN PERFORMANCE STABILITY

Robert E Schlegel
Randa L Shehab
School of Industrial Engineering

Kirby Gilliland
Department of Psychology

The University of Oklahoma
1000 Asp, Room 314
Norman, OK 73019-0631

Douglas R Eddy

NTI, Inc.
P.O. Box 35482
Brooks AFB, TX 78235-5104

Samuel G Schiflett

CREW SYSTEMS DIRECTORATE
Crew Technology Division
2504 Gillingham Drive, Suite 1
Brooks Air Force Base, TX 78235-5104

JULY 1995

Final Technical Report for Period January 1994 - June 1994

Approved for public release; distribution is unlimited.

AIR FORCE MATERIEL COMMAND
BROOKS AIR FORCE BASE, TEXAS

NOTICES

The voluntary, fully informed consent of the subjects used in this research was obtained as required by AFI 40-402.

This report has been reviewed and is approved for publication.

Sam Schiflett

SAMUEL G. SCHIFLETT, Ph.D.
Project Scientist

Patricia A Boll

PATRICIA A. BOLL
Acting Chief, Sustained Operations Branch

James P Dixon

JAMES P. DIXON, Colonel, USAF, BSC
Chief, Crew Technology Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
August 1995

3. REPORT TYPE AND DATES COVERED
Final Report May 92 - Oct 93

4. TITLE AND SUBTITLE
MICROGRAVITY EFFECTS ON COGNITIVE PERFORMANCE MEASURES: Practice Schedules to Acquire and Maintain Performance Stability

5. FUNDING NUMBERS
NASA T-6691R
F41624-91-C-2003
PE-62202F
PR-7930
TA-19
WU-15

6. AUTHOR(S)
Robert E. Schlegel, Randa L. Shehab, Kirby Gilliland, Douglas R. Eddy, Samuel G. Schiflett

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Oklahoma
School of Industrial Engineering
and Department of Psychology
1000 Asp, Room 314
Norman, OK 73019-0631

NTI, Inc.
P.O. Box 35482
Brooks AFB, TX 78235

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Armstrong Laboratory (AFMC)
Crew Systems Directorate
2504 Gillingham Dr, Ste 25
Brooks Air Force Base, TX 78235

National Aeronautics
and Space Admin.
Johnson Space Center
Houston, TX 77058

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AL/CF-TR-1994-0040

11. SUPPLEMENTARY NOTES
Armstrong Laboratory Technical Monitor: Dr Samuel G. Schiflett, (210) 536-3464

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
NASA is conducting a series of space shuttle launches to enable scientists to study the effects of microgravity. The Sustained Operations Branch of the USAF Armstrong Laboratory (AL/CFTO) has primary responsibility for studying the effects of microgravity on astronaut cognitive performance ability. To accurately identify performance decrements caused by microgravity in space, it is essential to collect preflight baseline data. Two studies were conducted to determine the impact on baseline performance stability of less than optimal practice schedules. In the first study, 21 subjects at Brooks AFB were trained on the NASA Performance Assessment Workstation (PAWS) and then assigned to one of five practice schedules. The study confirmed the overriding importance of providing an adequate number of practice sessions to achieve performance stability. Practice schedule interruptions had little impact on ultimate performance at the end of practice, provided the total number of sessions was maintained. An additional 80 subjects were tested using additional testing schedule alternatives. With few exceptions, the data showed remarkable consistency across the two studies. The data from both studies confirmed the high differential stability and reliability for the task measures and provided evidence for high software reliability. A database has been generated for classifying astronaut performance.

14. SUBJECT TERMS
Performance, Microgravity, Astronauts, Cognitive, Training

15. NUMBER OF PAGES
174

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

TABLE OF CONTENTS

ABSTRACT

NASA is conducting a series of space shuttle launches to enable scientists to study the effect of microgravity. The Sustained Operations Branch of the USAF Armstrong Laboratory (AL/CFTO) has primary responsibility for studying the effects of microgravity on astronaut cognitive performance ability. To accurately identify performance decrements caused by microgravity in space, it is essential to collect stable preflight baseline data. Two studies were conducted to determine the impact on baseline performance stability of less than optimal practice schedules. In the first study, 21 subjects at Brooks AFB were trained on the NASA Performance Assessment Workstation (PAWS) and then assigned to one of five practice schedules. The study confirmed the overriding importance of providing an adequate number of practice sessions to achieve performance stability. Practice schedule interruptions had little impact on ultimate performance at the end of practice, provided the total number of sessions was maintained. An additional 80 subjects were tested using additional testing schedule alternatives. With few exceptions, the data showed remarkable consistency across the two studies. The data from both studies confirmed the high differential stability and reliability for the task measures and provided evidence for high software reliability. A database has been generated for classifying astronaut performance.

PREFACE

This report documents the work performed at the University of Oklahoma under the prime contractor, Systems Research Laboratories, Inc., for the Armstrong Laboratory, Sustained Operations Branch (AL/CFTO), Brooks AFB under contract F41624-91-C-2003. Funding for the effort was provided by the National Aeronautics and Space Administration (NASA), through the Life Sciences Project Division at Johnson Space Center. A pilot study which provided data included in the performance database analyzed here was conducted as part of the U.S. Air Force Office of Scientific Research (AFOSR) Summer Faculty Research program administered by Research and Development Laboratories, Inc.

Several individuals deserve recognition for their contributions. Randa L. Shehab participated in the AFOSR Graduate Student Summer Research Program and served as project coordinator for the main study. Her contributions to the study design, subject recruitment and retention, data collection, reduction and analysis, and report writing were invaluable. The authors gratefully acknowledge the contributions of graduate research assistants Ioannis Vasmatzidis, Patrick L. Foster, and Tammy Kasbaum in the collection, reduction, summarization and analysis of the vast amounts of data. Special thanks are extended to Dr. William F. Storm for his friendship and genuine interest in the project. The authors wish to acknowledge the skilled and timely programming contributions of Kathy M. Winter, Sam J. LaCour, and Kathy Raynsford from the Naval Computer and Telecommunications Station, and Sam "Major" Moise, NTI, Inc. At several points during the research period, these individuals worked under tight deadlines to provide software changes needed to conduct the study. Software deadlines could not have been met without the early technology transfer actions of Dr. Frederick Hegge and Dr. Timothy Elsmore at the Walter Reed Army Institute of Research (WRAIR). Last, but certainly not least, enormous thanks are offered to those who really made the study possible - the experimental subjects. Their contribution of time and steadfast adherence to testing schedules (including weekends) provided the valuable data that enabled the project to be successful.

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

MICROGRAVITY EFFECTS ON COGNITIVE PERFORMANCE MEASURES:

PRACTICE SCHEDULES TO ACQUIRE AND MAINTAIN PERFORMANCE STABILITY

1.0 INTRODUCTION

The United States National Aeronautics and Space Administration (NASA) is conducting a series of space shuttle launches to enable scientists to study the effects of microgravity (weightlessness or near weightlessness) on a variety of factors. Included in the second International Microgravity Laboratory (IML-2) mission with a launch date of July 1994 is an extensive study of the effects of microgravity on astronaut cognitive performance. The Sustained Operations Branch of the USAF Armstrong Laboratory (AL/CFTO) has primary responsibility for this effort under the direction of Principal Investigator Dr. Samuel G. Schiflett (AL) and Associate Investigator Dr. Douglas R. Eddy from NTI, Inc. This large collaborative study includes astronaut training and testing on a battery of human performance tests prior to launch, periodically during the space mission, and after the flight. In preparing for this study, considerable research was conducted to establish the exact nature of the astronaut training and testing protocol and the degree to which the testing protocol would be robust with respect to such factors as launch delays, disruptions in the testing schedule, and various test parameters.

To permit an accurate identification of performance decrements caused by microgravity in space, it is essential to collect stable preflight baseline data. However, the astronaut's preflight schedule is extremely demanding. Only limited time is available for preflight training on any one project. Thus, there was a critical need to identify the optimal practice schedule, test parameters, and performance measures in order to achieve stable baseline performance data prior to the IML-2 space flight mission. It was also necessary to determine the impact of less than optimal practice schedules on the stability of the baseline performance data, and the influence of testing lapses (due to such factors as launch delays) on early in-flight performance. It was essential to understand the influence of such delays on baseline performance to determine when and how much additional practice may be needed after such delays to reestablish adequate baseline performance.

A preliminary examination of these issues was conducted during Summer 1992 under the U.S. Air Force Office of Scientific Research (AFOSR) Faculty and Graduate Student Summer Research Programs. This pilot study, which involved 21 male and female subjects with ages ranging from 23 to 52 years, provided a tentative answer to the question of optimal practice schedules and the impact of launch delays. The subjects were trained on the NASA Performance Assessment Battery and then assigned to one of five practice schedules. Two groups practiced each day for 15 consecutive days. Two other groups followed a schedule of 5 days testing, 2 days off, 5 days testing, 3 days off, 5 days testing. The fifth group followed a schedule of 2 days testing, 5 days off, 2 days testing, 5 days off, 2 days testing. Then, either 3 or 5 days after the last practice session, subjects returned for 5 days of retesting to represent mission days.

The study confirmed the overriding importance of providing an adequate number of practice sessions to achieve performance stability. By comparison, practice schedule interruptions (e.g., the 5 days on/2 days off schedule) had little impact on ultimate performance at the end of practice, provided the total number of sessions was maintained. High levels of differential stability and reliability were observed for at least one measure on all tasks except the Critical Tracking task. Excellent software reliability was demonstrated by less than 0.02% missed data collection points.

The larger study reported here was an expansion of the pilot study designed to examine additional combinations of practice schedules, testing lapses, a different measure of tracking performance, and flight test schedules using a substantially larger subject sample (80 subjects ranging in age from 17 to 44). It provided an opportunity for further refinement, testing and evaluation of various revisions of the task software, and created a large performance database for classifying astronaut performance.

This report summarizes the experimental design and methods used to address the outlined research questions. It provides graphical and statistical summaries and analyses of the performance data. Specifically, Section 2.0 presents the motivation for the project as well as specific research goals. Section 3.0 provides an extensive overview of the methodology and procedures used in this project, including a discussion of the various testing schedules, the subject characteristics, and the Performance Assessment Workstation (PAWS). This section is followed by Section 4.0 which presents the analyses of the data in terms of the specific research questions addressed. Section 4.0 also presents the overall performance database and a discussion of subject debriefing. Finally, Section 5.0 discusses the major research findings of the project. Extensive appendixes that provide additional detailed information about the project data complete the report.

2.0 OBJECTIVES

The Performance Assessment Workstation (PAWS) was assembled to investigate the effects of a space environment on human cognitive performance. The overall purpose of the research reported here was to determine the optimal PAWS testing schedule for astronauts to achieve stable baseline performance prior to flight and to determine the potential impact on flight data of test schedule interruptions resulting from launch delays and unforeseen events. Specific objectives were to:

(1) determine the optimal practice schedule for acquiring stable and reliable performance on the PAWS,

(2) determine whether the length of time lapse between practice and flight affects the retention of performance capability,

(3) demonstrate the stability and reliability of the test measures,

(4) evaluate and revise task parameters and levels for those tests not achieving stability or reliability,

(5) evaluate, improve, and verify the reliability of the testing software, and

(6) provide a larger performance database for classifying astronaut performance.

It might be hypothesized that missed testing days would result in degraded performance on one or more of the PAWS tasks and that a greater number of missed days would result in greater performance degradation. Subjects receiving continuous days of testing might be expected to achieve higher levels of performance and higher stability and reliability. Subjects undergoing abbreviated or discontinuous testing schedules might be expected to demonstrate degraded performance that would also be less stable and less reliable.

3.0 METHODOLOGY

3.1 Project Design

This project was designed to maximize the amount of information on practice schedule and testing lapse effects, given the available number of subjects. Although an infinite number of combinations of practice schedules, testing lapses, and mission test schedules could exist, the schedules used in this study were selected after careful consideration of probable scenarios for astronaut availability. The schedules are more easily discussed after an examination of Figure 1. In this figure, the "scheduled" launch date is the reference point for defining the schedules. Twelve different groups of subjects were tested. The number of subjects in each group is given in parentheses next to the Group Letter. With the exception of two groups tested under identical schedules (A1 and A2), the groups represent unique combinations of three scheduling factors: practice schedule, testing lapse between end of practice and first "mission" day, and mission test schedule. Mission refers to testing following practice to represent collecting inflight data.

Five different practice schedules were incorporated in the various groups. The nomenclature used to represent the groups in various tables and figures in this report is given below along with definitions of the schedules:

(1) **15 ED**: 15 consecutive days - one trial per test session (Groups A1, B, and A2),

(2) **5 on/2 off**: 5 days testing, 2 days off, 5 days testing, 3 days off, 5 days testing - one trial per test session (Groups C and D),

(3) **2 on/5 off**: 2 days testing, 5 days off, 2 days testing, 5 days off, 2 days testing - one trial per test session (Group E),

(4) **7 × 2 per day**: 8 consecutive days - two trials per test session except first day (Groups F, G, J, and K), and

(5) **15 × 2 (EOD)**: Every other day over 15 days - two trials per test session (Groups H and I).

Two different testing lapses following practice were examined, a 3-day lapse (Groups A1, C, A2, G, I, and K) and a 5-day lapse (Groups B, D, E, F, H, and J). Two mission testing schedules were used, one trial per session for 5 consecutive days (Groups A1 through G) and two trials per session every other day across 5 days (Groups H through K). Assignment to a particular schedule group was done after subjects completed a series of eight training trials. An

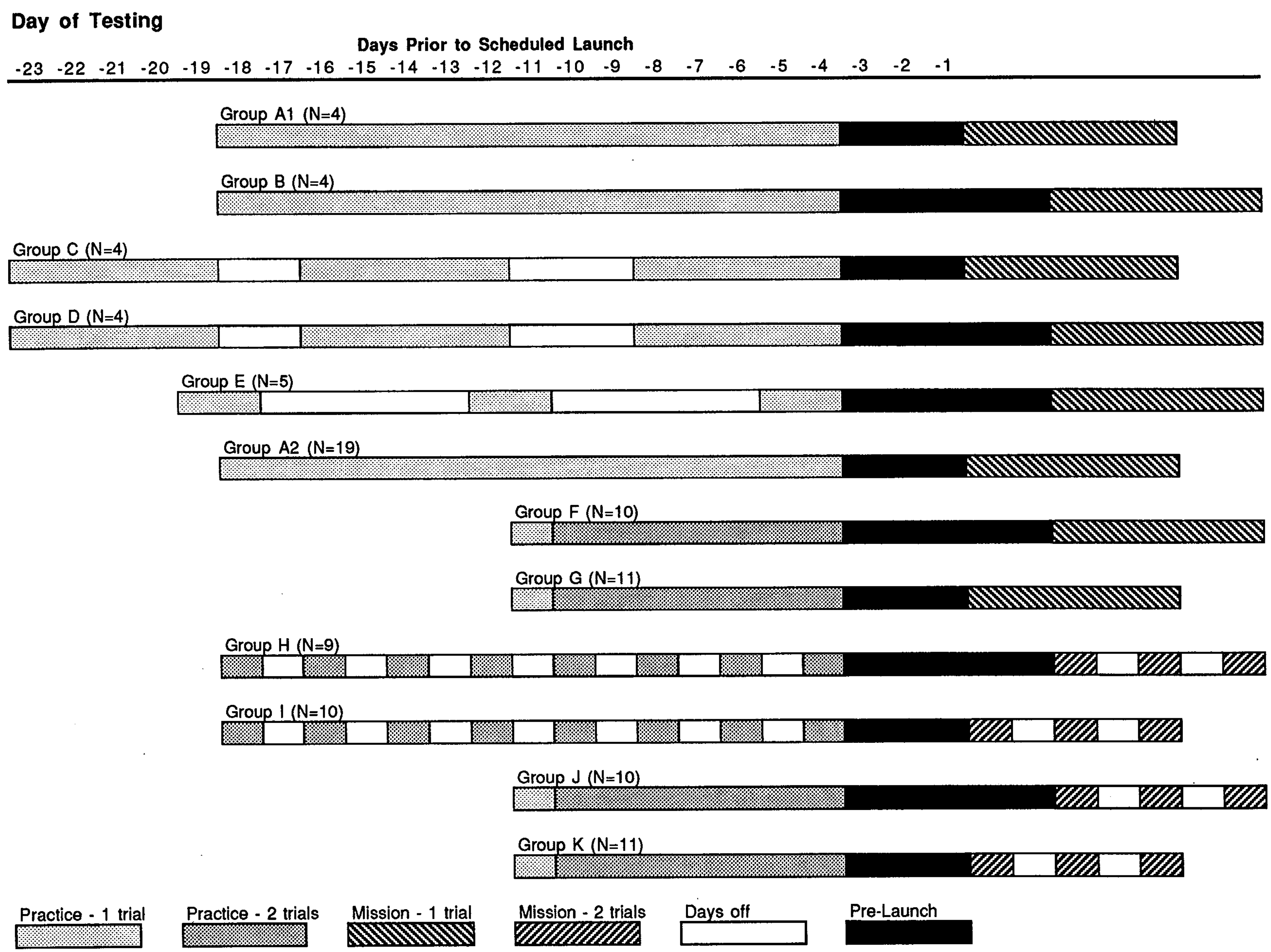


Figure 1. PAWS Testing Schedules.

attempt was made to balance the groups based on performance on the last training trial. The testing of Groups A1, B, C, D, and E (21 subjects) was completed during Summer 1992 at Brooks AFB. Groups A2, F, and G (39 subjects) completed testing during Fall 1992 and Groups H through K completed testing during Spring 1993 at the University of Oklahoma (OU).

3.2 Subjects

Subjects were recruited from two distinct populations. Data collected at Brooks AFB in Summer 1992 were obtained from subjects employed by the U.S. Air Force (USAF) as military or civilian workers or by various government contractors. Six women and 15 men ranging in age from 23 to 52 years with a mean of 35.1 and a standard deviation of 10.2 comprised Groups A1, B, C, D, and E. These subjects read and signed a voluntary consent form that was approved by the Brooks AFB Advisory Committee on Human Experimentation (ACHE). All summer subjects were volunteers and received no monetary compensation for their participation.

Subjects in the remaining groups, tested in Fall 1992 and Spring 1993, were recruited from University of Oklahoma psychology and engineering classes, the general student body, and the Norman, Oklahoma community. Of the 80 subjects recruited, 30 subjects were women and 50 subjects were men. They ranged in age from 17 to 44 years with a mean of 23.7 and a standard deviation of 5.3. These subjects signed an Informed Consent Form approved by the University of Oklahoma Institutional Review Board - Norman Campus (IRB-NC) which approved the experimental procedures in accordance with AFR 169-3. University of Oklahoma subjects were paid for their participation. Because this was a multi-session experiment involving testing on weekends, a bonus system was used to increase motivation and the completion rate. Subjects were paid $6.00 per hour for approximately 12.5 hours of testing ($75) and earned a $45.00 bonus for completing all trials at the scheduled times (total of $120). Only one subject was dropped part way through the study for lack of schedule compliance. Three other subjects retained their bonus money, but were not paid for one session that was missed.

Overall, 36 women (mean age of 23.4) and 65 men (mean age of 27.5) participated in the study. Ages ranged from 17 to 52 with an overall mean of 26.0 and a standard deviation of 8.1. All subjects were screened for self-reported normal (or corrected-to-normal) vision, normal hearing, and the absence of any central nervous system stimulant or depressant medications. A detailed breakdown of subject age characteristics for all subjects and for females and males is provided in Appendix A. The most obvious distinction is the age difference between the Brooks and OU samples. Differences in age among the various groups are merely a reflection of this general testing site difference. Of particular interest is a comparison of the performance groups A1 and A2 who used the same practice schedule, but differed in terms of age and testing site.

3.3 Test Battery

Several factors were considered in selecting tests for the NASA IML-2 experiment. One of the most important is the time available during flight for performance assessment. Another critical factor is the specific information processing skills necessary for mission success. The final and most relevant issue is the information provided by a specific test that could aid in identifying the cognitive processes or information processing stages affected by the space environment. These and other factors were taken into account in reviewing a large number of human performance task batteries. As a result, six performance tests were selected from the Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB). Also, two subjective scales were selected for inclusion in the NASA Performance Assessment Battery (NASA-PAB).

Brief descriptions of the performance tests and subjective rating scales selected for the IML-2 experiment follow in the order in which the tests are presented in the battery. Detailed descriptions of the tasks and a list of relevant publications are provided in Appendix B.

Mood Scale II - consists of 36 questions and involves pressing a numbered key to indicate the level of agreement with a descriptive adjective. The test takes 1 minute.

Critical Tracking - involves tracking an unstable object on the display using a trackball for 2 minutes.

Spatial Matrix - involves indicating whether a matrix of squares is the same as one previously presented. The test lasts 1.5 minutes.

Sternberg Memory Search - involves indicating whether a letter is the same as one of those in a previously memorized set. The test lasts 2 minutes.

Continuous Recognition Memory - involves pressing a key to indicate whether a number is the same as one previously memorized. The test lasts 2 minutes.

Switching Task - involves responding to 1 of 2 tasks presented simultaneously on each screen display. In the Manikin task, the subject presses a key to indicate which hand of a manikin holds a matching symbol. In the Mathematical Processing task, the subject presses a key to indicate whether a sum of three numbers is greater or less than 5. The test lasts 4 minutes.

Dual Task - involves performing the Sternberg Memory Search while Tracking and lasts 3 minutes.

Fatigue Scale - involves pressing a numbered key to indicate which statement best matches the subject's fatigue state and takes less than 15 seconds.

3.4 Equipment

The NASA-PAB is presented on a small portable microcomputer system. The hardware and software are collectively referred to as the Performance Assessment Workstation or PAWS. The PAWS consists of a GRiD 1530 laptop computer with an electroluminescent (EL) display and a NASA-compatible trackball. The GRiD uses a 32-bit Intel 80C386 central processor, has 4 Mb of RAM, an 80887 math co-processor, an internal 30 Mb hard drive, an internal 1.44 Mb floppy drive, an RS-232 interface, a built-in flat screen EL display, and a built-in keyboard. It uses GRiD MS-DOS (Microsoft Disk Operating System) and has a special ROM BIOS chip to allow MS-DOS to support the EL display screen. The PAWS requires a Measurement Systems, Inc. (MSI) 2-inch trackball (Model 622) connected to the serial port at 9600 baud and powered by 5 volts DC from a 120 volt AC/DC power supply.

3.5 Test Facilities

For the first 21 subjects, the overwhelming majority of testing sessions was conducted at subject workstations located in Room 24E of Bldg. 170 at Brooks AFB. Weekend testing sessions were conducted at subjects' residences. All testing of University of Oklahoma subjects was conducted in a three-room suite of laboratory space located in the basement of Dale Hall at the University of Oklahoma. Three PAWS testing stations approximately 3.0 ft. wide and 3.0 ft. deep were located in one room (approximately 13 ft. by 20 ft.). The stations were divided by 3-in. thick acoustic panels. The GRiD computers were placed on tables at the testing stations approximately 28 in. high.

Another room of approximately the same size served as a data reduction and project management office. The third room was used for interviewing, orientation, and miscellaneous activities. All of these rooms represent modern laboratory space with centrally controlled heating and air conditioning. Temperature in the testing room was maintained at approximately 68° F throughout the sessions.

3.6 Experimental Procedure

Data were collected from each subject over a period ranging from 6 to 8 weeks based on the assigned schedule. An orientation session provided subjects with background information about the project and instructions on the individual tasks. All subjects completed eight training trials. Brooks AFB subjects completed those trials in four 1-hour training sessions over a 2-day period (one session each morning and afternoon). University of Oklahoma subjects completed the training in four 1-hour sessions over a 4-day period (once per day at the same time of day). Subsequent practice and test sessions lasted 20 minutes per day for subjects on one-a-day schedules and 40 minutes per day for subjects on two-a-day schedules.

After the initial training routine, subjects were assigned to a specific schedule group, distinguished by practice schedule, length of testing lapse prior to "mission" testing, and mission testing schedule as defined in Section 3.1. Group assignment was based on two criteria. First, the highest/lowest performing subjects were distributed among the groups as evenly as possible. The remaining unassigned subjects were then placed in groups in order to minimize the differences of the average rank ordering of each group. Total testing time per subject for the entire study ranged from 12 to 14 hours.

For each trial, subjects performed the six tests and completed the two subjective scales using the PAWS for a total of 20 minutes. The tasks involved viewing a computer display screen, responding by pressing keys on the keyboard, and moving a trackball. The PAWS was designed to allow a subject to perform the tests independently without experimenter assistance. It is automated to minimize the time required for a well-practiced subject to perform the tests. It automatically performs all housekeeping functions, such as subject identification, file naming, test sequencing, and data backup. This feature was important for testing over weekends where some subjects administered the tests themselves without experimenter supervision.

In the initial training sessions, the interval between tests was subject-determined, that is, the tests did not start automatically. Subjects were required to press a key to start the next task. The subjects were allowed to ask questions and received feedback between tests. Summary feedback was provided at the end of each task during all sessions.

With the exception of the Dual Task, all task parameters remained constant throughout the experiment. Initially, the tracking component of the Dual Task was non adaptive (constant lambda of 2.0 for all subjects). However, the lambda value was individualized for each subject at a specific point during practice in order to equate task difficulty among subjects with different tracking abilities. The lambda value for each subject was fixed at 70% of the mean of the Critical Tracking maximum lambda values for a given number of trials. For subjects in Groups A1, B, C, and D, lambda was adjusted on Trial 15 based on the mean from the first six practice trials (Trials 9 through 14). For subjects in Group E, lambda was based on the mean of the first three practice trials (9 through 11) and was changed on Trial 12. Lambdas for the remaining subject groups were changed at earlier trials based on the observation that performance had not fully stabilized before mission testing for the previous groups. The lambda was based on the mean of the maximum lambdas obtained for Trials 7 and 8. Lambdas for Groups A2, F, and G were reset at Trial 10, for Groups H and I at Trial 11, and for Groups J and K at Trial 12.

4.0 RESULTS

This section of the report presents the data and analyses from this project beginning with descriptions of the magnitude of the data collection effort and the data reduction procedure. Data addressing the specific research questions are presented in the sections that follow. The presentation format is of a summary nature that should be particularly useful to those researchers interested in normative response patterns for each task. Also included for the convenience of the reader are tables with selected percentile groupings and box-and-whisker plots which allow classification of subjects into performance categories. Graphs for key performance measures and other detailed information are included in numerous appendixes. Following an initial discussion of the training data are sections that present the results of analyses of practice schedules, testing lapses, and mission schedules.

4.1 Data Reduction

This project involved the collection of a massive database. Only a portion of those data are summarized within this report. Table 1 presents a list of 28 key performance measures and the codes used to represent them in various tables and graphs. Almost 23,000 data observations (Subjects $\times$ Trials $\times$ Tasks), each containing numerous dependent measures, were collected over the course of 2377 subject sessions. It is noteworthy that of the 23,000 observations, less than 35 were lost due to equipment or procedural errors. Very few outlier datapoints were removed prior to the summaries and analyses. The deleted observations were due to identifiable subject errors. In instances where subjects inadvertently reversed response keys for an entire trial, the raw data files were rescored to provide correct summary information. Only 3 missing observations due to equipment or procedural problems and 10 observations deleted as outliers existed in the subset of data (5656 observations) used for the analysis of schedule effects.

The procedure for data reduction involved several phases. Generally, data reduction was performed after each of the Summer 1992, Fall 1992, and Spring 1993 data collection periods. For each of these large subject subgroups, the reduction followed the order in which data were collected; training data were reduced first, followed by practice data and mission data. Raw and summary data files from the individual subject PC diskettes were converted to files on an Apple Macintosh computer. Microsoft Excel macros were used to extract training data from the summary and/or raw files and to create individual subject graphs for the key performance measures of each task. These graphs were reviewed for questionable datapoints which could be

Table 1. Tasks and Performance Measures.

Task	Task Code	Measure Code	Description
Critical Tracking	TRK	LM ML CL RM	Lambda (maximum) Mean Lambda (mean of maximums) Number of Control Losses Root Mean Square (RMS) Error
Spatial Matrix	MTX	RT PC TP	Mean RT for Correct Responses Percentage of Correct Responses Throughput
Sternberg (Memory Search)	STN	RT PC TP	Mean RT for Correct Responses Percentage of Correct Responses Throughput
Continuous Recognition	CRC	RT PC TP	Mean RT for Correct Responses Percentage of Correct Responses Throughput
Switching Manikin Task	MAN	RT PC RTX PCX	Mean RT for Correct Responses Percentage of Correct Responses Mean RT for Transition Trials PC for Transition Time
Math Processing	MTH	RT PC RTX PCX	Mean RT for Correct Responses Percentage of Correct Responses Mean RT for Transition Trials PC for Transition Time
Dual Task	DUL	CL RM RT PC TP	Number of Control Losses Root Mean Square (RMS) Error Mean RT for Correct Responses Percentage of Correct Responses Throughput

associated with procedural errors or data outliers. Datapoints in question were corrected where possible, and removed when necessary. After verification that all of the training data were valid, this procedure was repeated for the practice data and the mission data. The next step in data reduction involved computing the means of the last two training trials and those of the last two practice trials. These data, combined with the mission data, were then plotted to obtain a picture of the general performance of each subject. A separate database for each task was assembled for the collection of subjects in each of the large subgroups.

4.2 Training Progress

To examine the pattern of skill acquisition for the various tasks, data for the eight training trials were summarized graphically. Due to minor differences in subject characteristics and testing protocol between the Summer, Fall and Spring subgroups, data were summarized separately for each of the three subgroups. As pointed out earlier, the Summer group containing a larger number of older individuals (mean of 35.1 vs. 23.7) was trained in eight sessions over a 2-day period rather than a 4-day period as with the Fall and Spring groups. For the Spring group, a greater emphasis was placed on initial speed for the Matrix task. Subjects were told that they may do better if they did not try to memorize the exact locations of all filled cells, but looked at the pattern as a whole and moved rapidly from one pattern to the next.

Charts summarizing the training performance of all subjects collapsed across testing periods (Summer, Fall, and Spring) are provided in Figures 2 through 5 and in Appendix C. Task learning is indicated by faster response times, higher accuracy and throughput, and fewer control losses over time. In general, performance improved rapidly over the first three to five trials. The rate of improvement leveled off by the eighth trial. Similarities and differences between the groups are pointed out on a task-by-task basis in the discussion that follows.

Mood Scale II

Before starting the performance tasks, all participants reported their moods by responding to adjectives using a 3-point scale. A response of "1" indicated that the subject did not feel that the adjective described the current mood while a response of "3" indicated that the adjective adequately described the subject's mood. The adjectives are divided into six categories (Activity, Happiness, Depression, Anger, Fatigue, and Fear). Regardless of the study period, all subjects on the average ranked high in Activity and Happiness. For both of these categories, the mean scores were around 2.5 on the 3-point scale for all eight trials. On the contrary, the scores in the remaining four categories (Depression, Anger, Fatigue and Fear) were very close to one (the lowest possible score). The Fatigue category scores were slightly higher and occasionally reached values between 1.3 and 1.5. In general, the Brooks AFB subjects (Summer 1992) were slightly more active and happier, and had lower scores on the "negative mood" categories.

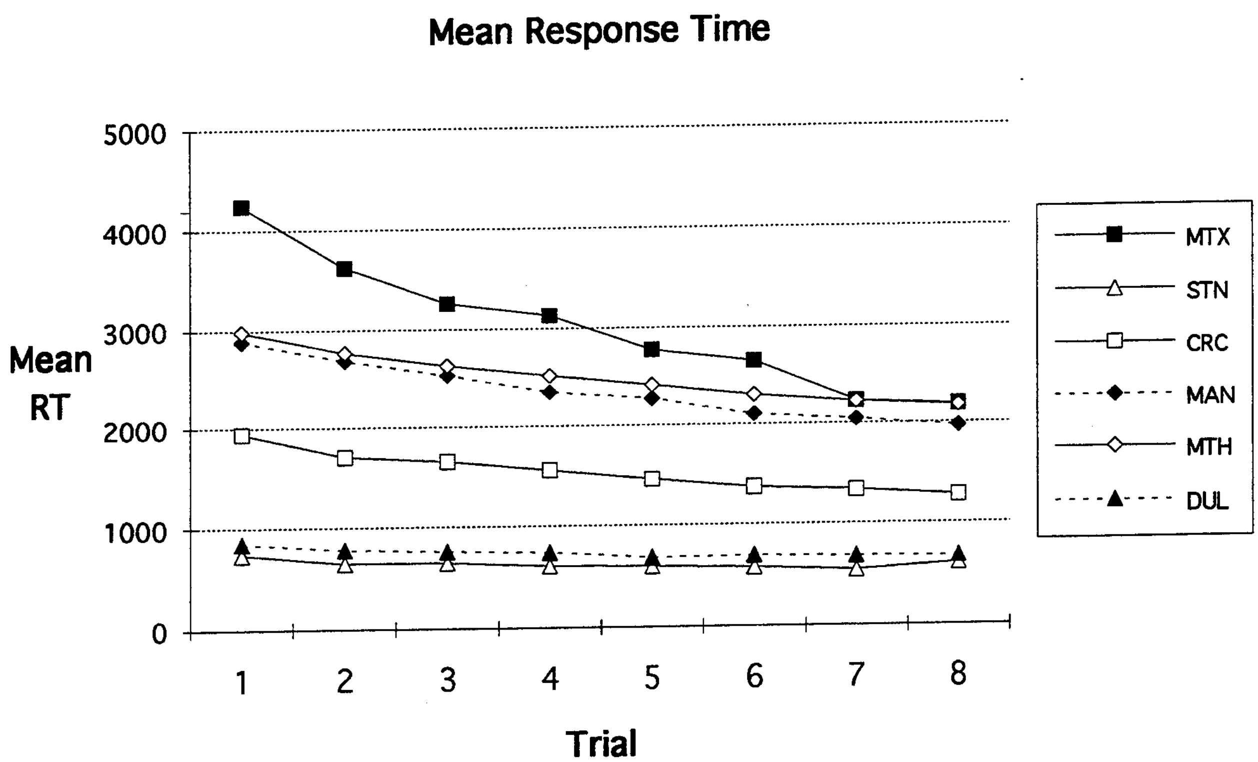


Figure 2. Mean Response Time for Discrete Tasks - Training Sessions.

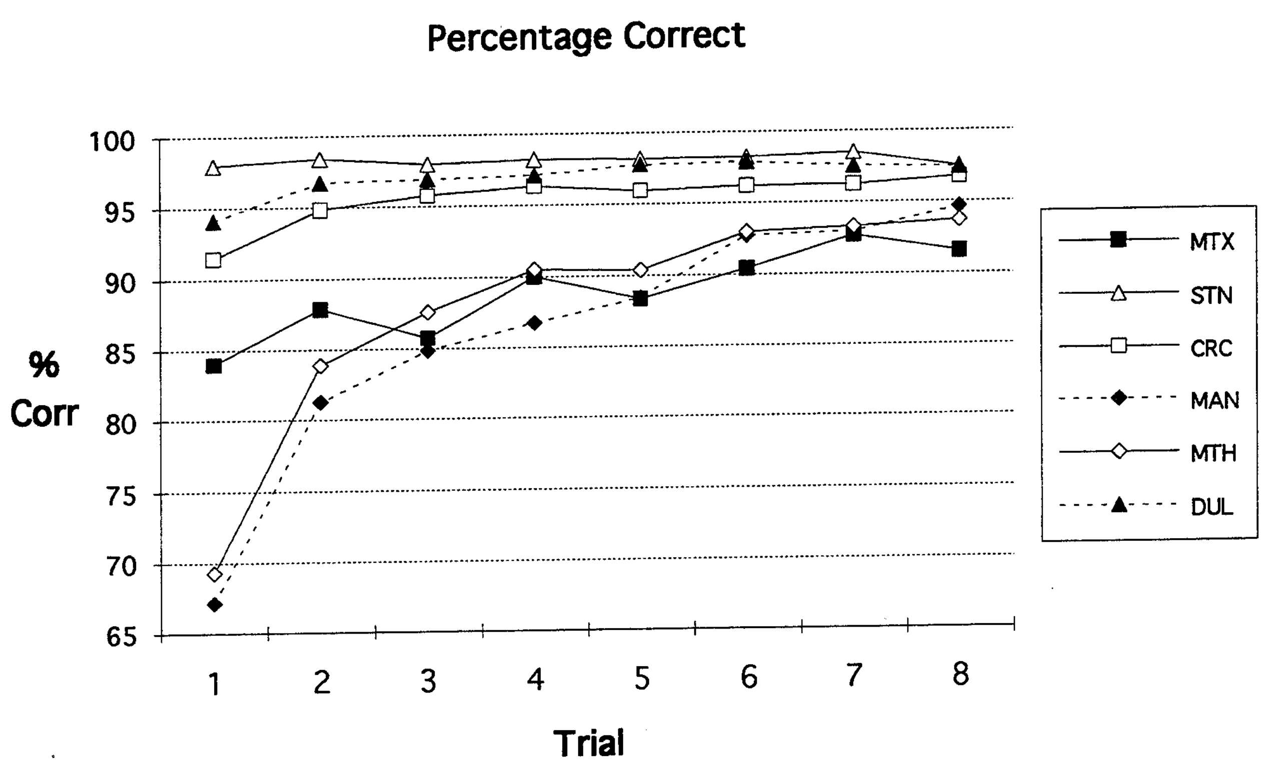


Figure 3. Percentage Correct for Discrete Tasks - Training Sessions.

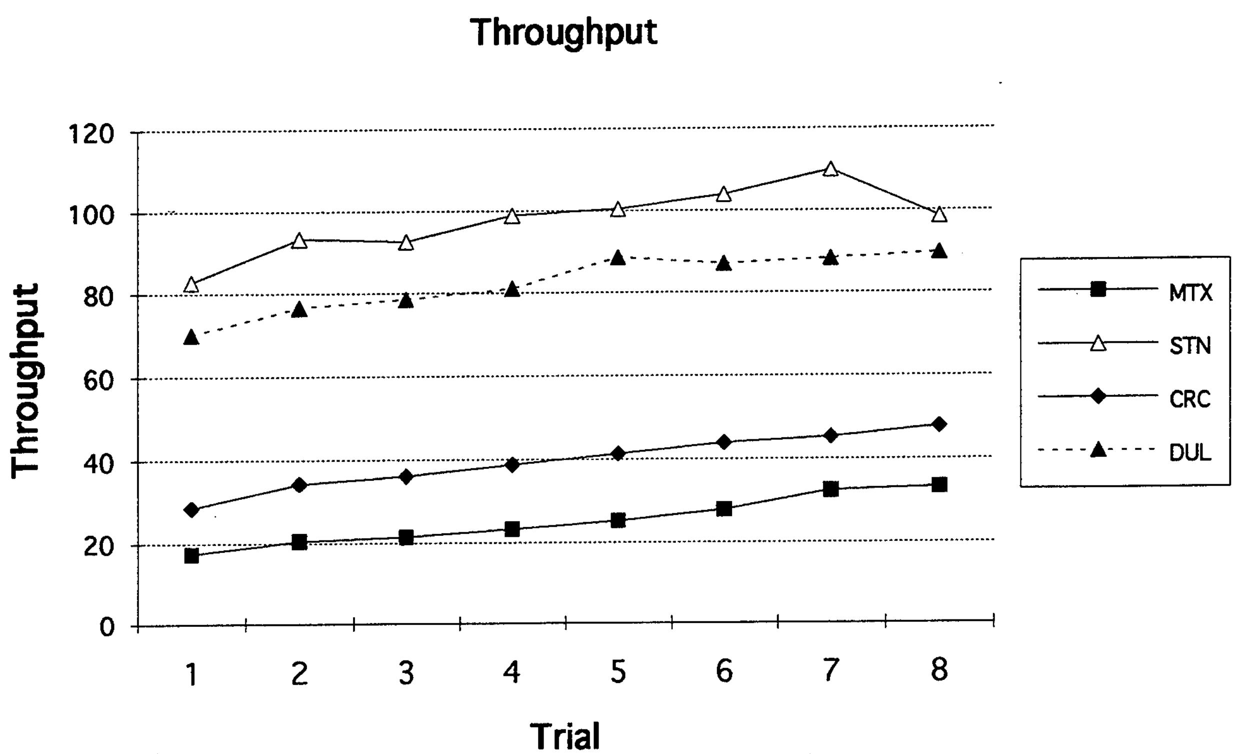


Figure 4. Throughput for Discrete Tasks - Training Sessions.

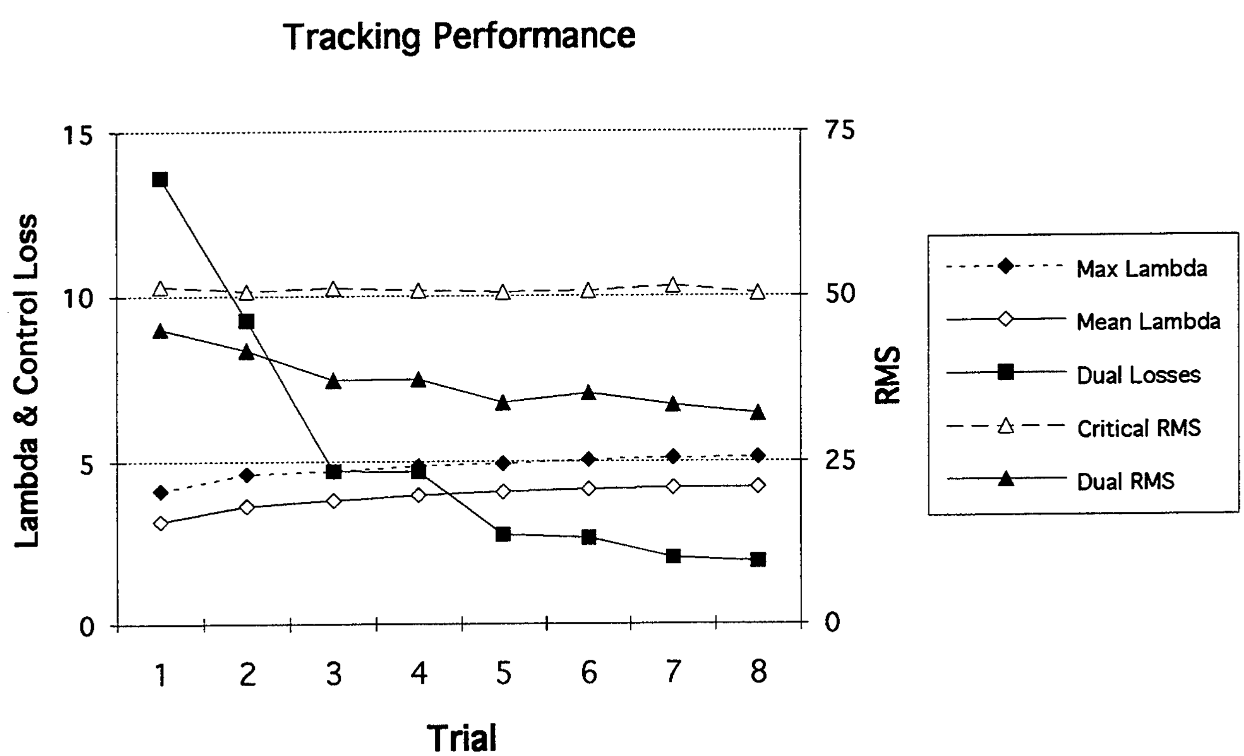


Figure 5. Performance Data for Tracking Tasks - Training Sessions.

With respect to response times, it took the subjects a little longer to respond to the Activity and Happiness adjectives. There are two possible explanations for this. First, it is possible that people with positive attitudes such as those in this study can quickly decide that they are not depressed, angry, etc., but may take longer to determine the extent of their activity and happiness levels. Another explanation relates to the fact that since there are roughly twice as many "negative" adjectives (four categories) as "positive" adjectives (two categories), there is priming on the part of the subjects to respond using the "1" key (corresponding to the "NO, not at all" response) for the "negative" adjectives. There would then be a slight delay when the less frequent "positive mood" adjectives appeared. The time to respond to the Fatigue category was somewhat longer than for the other three categories describing a "negative mood", perhaps indicating that more thought was required to respond to these adjectives. Overall, response times decreased from the first to the last training trial for all three testing periods.

Critical Tracking

Critical tracking performance was assessed in terms of the maximum lambda during the trial, the mean of the lambdas associated with each control loss during the trial (for the Fall and Spring subjects only), number of control losses, and Root Mean Square (RMS) error. Maximum lambda improved throughout training. Summer subjects improved from a maximum lambda of 4.6 on Trial 1 to approximately 5.0 for Trial 8. Fall and Spring subjects started lower (4.0 for Trial 1) but achieved a similar level of proficiency by the end of the training period. The biggest performance improvement in maximum lambda occurred between Trial 1 and Trial 2. The similarity in learning patterns between the Fall and Spring groups with respect to maximum lambda, was also reflected in the mean lambda measure (i.e., the measures tracked each other well). In particular, the average mean lambda for both groups ranged from about 3.0 for Trial 1 to just above 4.0 for Trial 8.

The number of control losses for the Critical Tracking task is inversely related to a subject's maximum lambda and serves only as a crude measure of tracking ability because it is very dependent on task parameters. For the Summer subjects, the average number of control losses decreased from 13 (Trial 1) to 11 (Trial 8). The respective changes for the Fall and Spring participants were from 17 (Trial 1) to 12 (Trial 8) and from 18 (Trial 1) to 12 (Trial 8). RMS error was not a sensitive measure of performance and remained essentially constant throughout training for all three study periods. When the task is run in "Critical" mode, this measure represents a performance summary across a wide range of tracking conditions (lambdas) and as such is not very indicative of tracking ability. In fact, as lambda increases, more trackball and cursor movement is required and RMS error must naturally increase. Both measures (number of control losses and RMS error) are more valid under fixed lambda conditions as in the Dual Task.

Matrix

Three measures were used to assess performance on the Matrix, Memory Search, and Continuous Recognition tasks: response time, percent correct and throughput. Matrix response times were similar for the Summer and Fall groups, decreasing from approximately 5000 msec in Trial 1 to 2500 msec and 3000 msec, respectively, in Trial 8. As stated previously, subjects in the Spring session were encouraged to view the patterns more rapidly rather than memorizing all cell locations. This strategy resulted in reduced mean response times from the start of training (3000 msec on Trial 1) as well as greater improvement by the end of training (1800 msec on Trial 8). It is believed that the difference in instructions, coupled with close monitoring of subject performance during the early trials, was responsible for the faster RTs which were obtained without sacrificing accuracy. Percentage correct improved across training from about 85% to about 92% for all three groups. The faster RTs of the Spring subjects are also evident in the throughput data. For both Summer and Fall subjects, throughput rose steadily from 15 to about 30. Throughput was consistently higher for the Spring subjects and ranged from a little over 20 (Trial 1) to almost 40 at the end of training.

Sternberg Memory Search

In general, response times for Memory Search became shorter as training progressed. However, this trend was not as prominent with the Summer subjects. These subjects improved from a mean RT of 700 msec on the first trial to a little above 600 msec by the end of the training. The corresponding improvements for both the Fall and Spring participants were from about 750 msec to below 600 msec. An interesting pattern in the response time results for all three groups is the increase in RT for Trial 8 following a minimum on Trial 7. This worsening of performance on the last training trial has been observed by the authors in several studies and seems to reflect a letdown in the subjects when they have reached what they know is the end of a study phase (e.g., end of training or end of practice).

Percent correct remained high throughout training and exhibited minimal fluctuation. In Summer, percent correct was very close to 97% for all trials. For Fall and Spring this measure essentially stabilized at 98%. The throughput patterns were similar for all three testing periods. Throughputs started at approximately 80 for Trial 1 and reached a peak in the region of 110 for Trial 7. As with mean RT, throughput worsened by ten or more units from Trial 7 to Trial 8.

Continuous Recognition

For Continuous Recognition, the similarities in performance among the groups were incredible. Response times for all three groups declined from an initial level of almost 2000 msec to about 1300-1400 msec by the end of training. In terms of average percent correct, an identical change was obtained for the Fall and Spring participants, starting at 92% and moving to 98%. Improvement for the Summer participants was not as dramatic, improving from 92% to 93% during training. Throughput improved steadily for all groups across the eight trials of training, moving from just below 30 to just above 40 for the Summer subjects and from just below 30 to just above 50 for the Fall and Spring subjects.

Switching

Performance on this task was assessed separately for the Manikin and the Mathematical Processing portions of the task. Manikin response times declined from 3000 msec to 2200 msec in Summer, and from 2800 msec to about 2000 msec in Fall and Spring. Response times for transition stimuli, obtained only for the Spring subjects, were slightly slower than the corresponding overall response times across the entire training period. This difference between transition and overall response times ranged from 40 to 270 msec. Manikin percent correct improved dramatically for all study periods. While the initial values for this measure were somewhere between 65% and 70%, the final values exceeded 90% and ranged from 92% to 94%. The slower performance on transition trials was not accompanied by a lower percent correct beyond the first two trials of the training period.

The pattern of response times for the Mathematical Processing task was similar to that of the Manikin task. In Summer, response times decreased from over 3000 msec to 2200 msec. In Fall and Spring the subjects started a little faster (just below 3000 msec), but reached the same final speed of 2200 msec. Once again, the transition trials were associated with slower response times; there was an average difference between transition and overall RTs of about 200 msec. Mathematical Processing percent correct also improved to a great extent; most of the improvement took place between Trials 1 and 2. This measure jumped from 65% to 92% for the Summer participants, from 72% to 93% for the Fall participants, and from 68% to 93% for the Spring participants. Moreover, transition percent correct in Spring was lower by about 2% than the overall percent correct values with the exception of Trial 3 in which the difference was larger (about 5%).

Dual Task

Performance on the Tracking portion of the Dual Task was assessed using the measures of control losses and RMS error. Both measures revealed that the Summer subjects initially performed better than the Fall and Spring subjects. In Summer, control losses decreased from 7 in Trial 1 to approximately 3 at the end of training (although a small increase in control losses was observed between Trials 7 and 8). For the Fall and Spring periods, the changes were from 13 to 2 and from 17 to 2 respectively. In Summer, RMS error improved slightly from the first trial to the second trial and then remained essentially constant at 30. However, the Fall and Spring subjects started higher (45 for Fall and 50 for Spring) but reached the same RMS of 30 by the end of training.

Performance on the Sternberg Memory Search portion of the Dual Task was assessed in terms of the usual measures of response time, percent correct, and throughput. The initial levels of response time were different for each group (780 msec in Summer, 820 msec in Fall, and 900 msec in Spring), but all groups converged to a similar speed of 680-690 msec on Trial 8. This pattern indicates that an RT floor value was probably reached by the end of training. The percent correct measure changed in similar fashion for all three groups. In Summer, percent correct improved from 94% to 97%; in Spring it improved from 95% to 97%; and in Fall the change was from 93% to 98%. A similar pattern of improvement among the three groups was observed for throughput which improved from 75 to 88 in Summer, from 72 to 88 in Fall, and from 65 to 92 in Spring.

Fatigue Scale

The subject fatigue level at the end of each experimental trial was captured using a seven-point scale (with "7" indicating maximal fatigue). In Summer, training consisted of four trials performed by the subjects on each day of a 2-day period. As expected, fatigue developed progressively over the four trials each day and this increased fatigue level was reflected in the subject responses. In particular, the fatigue level increased from Trial 1 to Trial 4 on the first day and from Trial 5 to Trial 8 on the second day. In the Spring and Fall (when training was conducted with two trials per day for 4 days), there was an increase in fatigue from the first to the second trial each day. Response times decreased from about 8 seconds (Fall) or 10 seconds (Spring) for Trial 1 to less than 5 seconds for Trial 8. A smaller decrease in response time was observed in the Summer.

4.3 Test Site Differences

The data tables in Appendix D illustrate the high level of agreement in the data collected at Brooks AFB and the University of Oklahoma. For each key performance measure, the data labeled "Training" represent the mean of the performance measure across Trials 7 and 8; "Start" represents Trial 9 (the first trial of practice after a break of approximately 2 weeks following training); "End" represents the mean of Trials 22 and 23 at the end of practice; and MD-1 through MD-5 represent the five trials simulating "Mission Day" testing. Due to the dramatic difference in the number of practice trials, the data for Group E are excluded from these summaries. No statistical analysis of test site differences was conducted. A brief discussion of observations based on the data for each task follows.

Critical Tracking

The average maximum lambda at the end of training was the same for both groups. During practice, Brooks subjects achieved a higher level of performance (6.1 vs. 5.7 for OU subjects). A smaller difference between groups persisted throughout mission testing. A similar pattern was observed with respect to control losses, which were approximately 1 unit lower for Brooks subjects by the end of practice. Interestingly, a constant difference between groups existed for RMS error across all trials (training, practice, and mission) with OU subjects having slightly higher RMS values.

Matrix

Large differences in Matrix response times existed initially between the subject groups due in large part to the difference in instructions for half of the OU subjects (Spring 1993). This difference was reduced substantially during practice (from 500 msec to 200 msec) and was virtually eliminated halfway through mission testing. Following training, there was essentially no difference in percent correct between the groups. Throughput differences reflected the differences in RT throughout practice, but no difference was observed during mission testing.

Memory Search

The mean RT for Memory Search at the end of training was the same for both groups. During practice, Brooks subjects achieved lower RTs (500 msec) than OU subjects (540 msec) and this 40 msec difference persisted throughout mission testing. Percent correct for the two groups differed slightly during training, became equivalent by the end of practice and separated again during mission testing. OU subjects were slightly more accurate on the average (98%

compared to 97% for Brooks subjects). The RT and percent correct differences tended to cancel in a way that throughput was essentially equivalent for the two groups throughout all testing.

Continuous Recognition

Although they were slightly different during training, the mean RTs for the two groups were essentially the same during practice and mission testing. Percent correct was a constant 5% lower for the Brooks subjects due primarily to the substantially lower accuracy of one Brooks subject group. Consequently, throughput was lower for the Brooks group at the end of training and throughout practice, but did not differ during mission testing.

Switching

For the Manikin task, the mean response time was slightly higher (200 msec) for the Brooks subjects at the end of training. By the end of practice, this difference had virtually disappeared. Mean RT was essentially equivalent for both groups during mission testing. The OU group maintained a consistently higher percent correct (by approximately 1.5%) throughout all testing.

Response times for the Mathematical Processing portion of the Switching task were equivalent for both groups. Percent correct for the OU group was lower by 2% at the start of practice, but exceeded the Brooks group at the end of practice by 1%. The OU group remained higher throughout mission testing by as much as 4%.

Dual Task

Although OU subjects started with fewer control losses on Dual-Tracking, they had substantially higher control losses (12) than Brooks subjects (7) by the end of practice. This difference persisted throughout mission testing, but fluctuated from as small a difference as 2 to as large a difference as 6. RMS error was consistently higher for OU subjects (by approximately 5 units) throughout all testing. For Dual-Memory Search, mean RT values were essentially equivalent and constant across the testing trials. These results can be misleading between the start and end of practice due to the change in lambda part way through practice. This is discussed in more detail in Section 4.4. Percent correct was constant for both groups but slightly lower for the Brooks subjects (approximately 2%). This minor difference in percent correct was washed out by the consistency in RT; consequently, throughput was identical for both groups.

Fatigue Scale

Across practice, fatigue scores remained relatively constant for OU subjects, but actually decreased for the Brooks subjects. Subjective fatigue appeared to decrease slightly across mission testing for the OU group while it randomly fluctuated around a constant level for the Brooks subjects.

4.4 Practice Schedules

Twelve different testing groups were examined during the course of the yearlong study. Two of the groups (A1 and A2) actually ran the same schedule (the "optimal" schedule of 15 consecutive days), but they represented different subject samples (Brooks vs. OU). The performance data are summarized by test group in Appendix E. As noted previously, the training data point is the mean of Trials 7 and 8, the Start of practice is Trial 9, and the End of practice is the mean of Trials 22 and 23 (Trials 13 and 14 for Group E). Although the data tables may be examined to identify the source of specific subject group differences, a clearer picture can be obtained by collapsing the twelve subgroups according to the various schedule characteristics. Appendix F presents graphs of the data by practice schedule, the primary factor of interest in the study.

Visually, there are very few apparent differences between the practice schedules with the exception of the 2 on/5 off schedule. For almost all task measures, average performance continued to improve throughout the practice period; the rate of improvement depended on the particular task and measure. In general, the rate of improvement was the same for all schedule groups. However, because Group E subjects had only six trials of practice, they did not, in general, achieve the same level of performance by the end of their relatively short practice period as the remaining groups with 15 practice trials. In the graphs, this outcome is recognized as a smaller slope since only start and end of practice are plotted (i.e., the time units are not equivalent for Group E).

Systat Version 5.1 was used to conduct repeated measures analyses of variance (ANOVA's) on the practice and mission data. Because the overall analysis included numerous tests, it would have been appropriate to adjust the experiment-wise Type I error rate through some procedure such as dividing alpha. Of course, in this case, where one might hope that no differences would emerge among the subject groups, Type I error rate control procedures make the detection of significant differences more difficult -- and thus, work in favor of finding fewer significant differences. A statistically less conservative approach was taken in the analyses for this project by not adjusting the Type I error rate. It was further determined that a large analysis involving training, practice and mission trials would wash out the statistical significance of any effects of interest. Thus, analyses involving only the "Start" of practice and "End" of practice data points, were conducted.

It was expected that there might be raw-score differences between the various groups merely due to random chance group assignments. It was also expected that most tasks would demonstrate continued performance improvement from start to end of practice. However, the key question was whether a differential amount of improvement would occur as a result of assignment to a particular practice schedule. This improvement difference would be evidenced by a different learning slope for those groups and could be tested statistically by examining the Group by Trial interaction. In general, subjects in all groups, with the exception of 2 on/5 off, followed an amazingly similar pattern of performance improvement on the average. Exceptions to this general statement are discussed below.

Table 2a summarizes the results of the ANOVA examining the differential impact of the five practice schedules on performance improvement from start to end of practice. In this analysis, Schedule is a between-subjects factor that is representative of raw-score differences between the groups. These differences may result naturally from the assignment of subjects to practice schedules, or may be due to differences in subject characteristics or training protocols. In examining Table 2a, it is in fact surprising that few such differences existed and those that did are easily explained. Three differences worthy of discussion are now presented.

TRK LM, CL - A marginal difference in Critical Tracking maximum lambda and number of control losses existed due to the poorer overall performance of the subjects following the every-other-day/two trials per day (15×2) practice schedule.

MTX RT - Response time on the Matrix task was significantly faster for the two-a-day schedules (7×2 and 15×2). These schedule groups were primarily composed of subjects who were tested in Spring 1993. As noted before, these subjects were encouraged to view the Matrix patterns in rapid succession rather than memorizing individual cell locations. The RT differences between the Summer, Fall, and Spring groups noted in the discussion on Training Progress (Section 4.2) persisted throughout practice and were identified by the ANOVA ($F(4,96) = 4.14$, $p < 0.01$). The five subjects in Group E (2 on/5 off) exhibited particularly long RTs throughout all testing sessions. These differences due to initial subject instructions and poor Group E performance were also reflected in a marginally significant difference in the throughput measure.

Table 2. ANOVA Results Comparing Start and End of Practice by Practice Schedule.

Results include Group E (Table 2a)

Task	Measure	Schedule (S) F(4,96)	Schedule (S) p	Trial (T) F(1,96)	Trial (T) p	S x T F(4,96)	S x T p
TRK	LM	2.24	0.07	86.94	**0.00**	4.89	0.00
	CL	2.22	0.07	97.76	**0.00**	1.51	0.21
	RM	1.31	0.27	0.00	0.96	0.64	0.64
MTX	RT	4.14	0.00	22.49	**0.00**	1.01	0.40
	PC	0.64	0.64	24.46	**0.00**	0.42	0.79
	TP	2.12	0.08	60.03	**0.00**	3.55	0.01
STN	RT	0.60	0.66	43.94	**0.00**	1.00	0.41
	PC	1.14	0.34	2.71	0.10	0.85	0.49
	TP	0.70	0.59	67.24	**0.00**	1.60	0.18
CRC	RT	0.25	0.91	134.09	**0.00**	2.38	0.06
	PC	3.85	0.01	0.32	0.58	0.13	0.97
	TP	0.29	0.89	177.00	**0.00**	4.93	0.00
MAN	RT	0.46	0.76	143.82	**0.00**	2.04	0.10
	PC	0.88	0.48	9.76	**0.00**	0.47	0.76
MTH	RT	0.36	0.84	86.18	**0.00**	1.69	0.16
	PC	0.82	0.51	16.65	**0.00**	1.22	0.31
DUL	CL	1.30	0.28	56.76	**0.00**	2.71	0.03
	RM	0.67	0.61	93.04	**0.00**	1.38	0.25
	RT	0.70	0.59	0.46	0.50	0.92	0.46
	PC	0.66	0.62	1.60	0.21	0.37	0.83
	TP	0.56	0.69	0.36	0.55	0.97	0.43
FAT	RP	1.41	0.24	1.78	0.19	2.39	0.06

Results omit Group E (Table 2b)

Task	Measure	Schedule (S) F(4,96)	Schedule (S) p	Trial (T) F(1,96)	Trial (T) p	S x T F(4,96)	S x T p
TRK	LM	3.02	0.03	178.23	**0.00**	2.24	0.09
	CL	2.94	0.04	152.58	**0.00**	0.68	0.56
	RM	1.24	0.30	0.01	0.92	0.90	0.45
MTX	RT	4.88	0.00	34.11	**0.00**	0.96	0.41
	PC	0.92	0.44	33.51	**0.00**	0.45	0.72
	TP	1.98	0.12	96.82	**0.00**	2.95	0.04
STN	RT	0.74	0.53	58.63	**0.00**	1.12	0.34
	PC	1.16	0.33	3.60	0.06	1.09	0.36
	TP	0.91	0.44	85.16	**0.00**	2.05	0.11
CRC	RT	0.04	0.99	192.37	**0.00**	2.70	0.05
	PC	4.94	0.00	0.40	0.53	0.20	0.90
	TP	0.20	0.90	272.54	**0.00**	3.78	0.01
MAN	RT	0.24	0.87	211.56	**0.00**	1.08	0.36
	PC	0.56	0.65	18.51	**0.00**	0.01	1.00
MTH	RT	0.27	0.85	122.68	**0.00**	1.16	0.33
	PC	0.54	0.66	25.33	**0.00**	1.35	0.26
DUL	CL	0.07	0.98	35.55	**0.00**	1.34	0.27
	RM	0.87	0.46	107.77	**0.00**	1.91	0.13
	RT	0.82	0.48	0.12	0.73	1.10	0.35
	PC	0.65	0.58	0.75	0.39	0.36	0.78
	TP	0.71	0.55	0.04	0.85	0.83	0.48
FAT	RP	1.78	0.16	1.17	0.28	3.23	0.03

CRC PC - Percent correct for Continuous Recognition was significantly lower for the subjects assigned to the 5 on/2 off schedule. This result is likely to have been due to the specific subjects assigned to this small group (n = 8).

It was expected that all performance measures would show improvement across the course of the practice sessions. This expectation was tested with the within-subjects factor Trial. With few exceptions, the differences between the start and end of practice were highly significant. The exceptions are noted below.

TRK RM - As noted in Section 4.2, RMS error for Critical Tracking is not a valid measure of tracking ability. Thus, there was no change in this measure across practice trials.

STN PC - Percent correct for Memory Search was uniformly high (98%) throughout practice.

CRC PC - Percent correct for Continuous Recognition was essentially constant at 91% for the 5 on/2 off schedule and 97% for the remaining subjects.

DUL RT, PC, TP - The apparent lack of performance improvement for the Sternberg portion of the Dual Task is misleading. Keep in mind that the difficulty of the Tracking task was increased for all groups at some point during practice. This difficulty change also had an impact on Sternberg performance which typically worsened for the first few trials following the lambda change and thereafter improved only to the original performance levels. In fact, some groups exhibited slightly worse Memory Search performance at the end of practice than at the start.

The true impact of the various practice schedules is identified by testing the Schedule × Trial interaction. This interaction represents the differential improvement of the five practice schedules from the start to the end of practice. Five tasks yielded significant interactions.

TRK LM - Tracking lambda improved at varying rates for all groups except Group E (2 on/5 off) where lambda actually decreased across practice. Some improvement for Group E occurred during mission testing.

MTX TP - The extent of improvement in Matrix throughput was much lower for Group E (2.7 units) compared with the mean improvement for the other groups (13.1 units).

CRC TP (RT marginal) - The extent of improvement in Continuous Recognition was much lower for Group E (4.7 units) compared with the mean improvement for the other groups (16.0 units).

DUL CL - With the exception of the 5 on/2 off group, all schedule groups started uniformly low with respect to control losses. Due to the small number of practice trials following the lambda change, Group E subjects had substantially higher control losses at the end of practice (25 vs. 11 for the other groups).

FAT RP - By the end of practice, the fatigue level was higher for the 7 day × 2 per day schedule (3.7) than for the other schedules (2.7). This difference may have been due to the fact that this schedule was unique. All practice trials were completed within 7 1/2 days compared with 15 or more days for all other schedule groups.

Group E (2 on/5 off) was unique. Only six trials were completed over a 20-day period. This placed the Group E subjects at a distinct disadvantage throughout the practice period and during mission testing, so much so that subjects in this group continued to improve on many of the tasks throughout the five mission trials. To determine the extent to which the ANOVA results were influenced by Group E performance, the analyses were repeated after excluding the data for this group. The results of these analyses are presented in Table 2b. With respect to raw-score differences between the schedule groups, the same significant measures were identified. The degree of significance for Critical Tracking lambda and control losses increased. The 15 × 2 (EOD) schedule demonstrated poorer overall performance than the remaining three schedules. Again, this performance difference existed during training and at the start and end of practice; it does not represent an effect caused by the practice schedule itself. With respect to the within-subjects Trial factor, there were no changes in the significant key measures with Group E removed.

With respect to the Schedule × Trial interaction, the previously significant Critical Tracking lambda and Dual Task control losses lost their significance, confirming that the previous effect was due primarily to Group E. The effect on Matrix throughput ($p = 0.04$) and Continuous Recognition throughput ($p = 0.01$) was also weakened while the effect on fatigue level was strengthened ($p = 0.03$) by the removal of Group E from the analysis.

MTX TP - Subjects assigned to the 15 × 2 (EOD) practice schedule did not increase their performance at as high a rate as subjects in the remaining three schedule groups (8.9 units vs. 14.5 units).

CRC TP - Although not as obvious from the graph, subjects in both two trial per day schedules improved their performance at a slightly higher rate than subjects in the other two schedules.

FAT RP - Removal of Group E fatigue data statistically strengthened the previously mentioned fatigue effect for the 7 × 2 group.

4.5 Testing Lapses

In order to evaluate the impact of testing lapses (3-day lapse vs. 5-day lapse) on performance, the data from all groups except Group E were collapsed to produce the graphs in Appendix G. A repeated measures analysis using Practice Schedule and Testing Lapse as between-subjects factors was conducted. Data from the end of practice and the first mission trial were included in the analyses and Trial served as a within-subjects factor. As in the analysis of practice schedule, a significant finding for Schedule or Lapse merely indicates an overall performance difference due to the random assignment of specific individuals to specific groups. The results of the ANOVAs are summarized in Table 3.

For the factor Schedule, only four measures attained significance. The charts in Appendix F help illustrate the differences if one visually averages the data from the end of practice and MD-1 for each schedule.

TRK LM, CL - The marginally significant difference in Critical Tracking maximum lambda as well as the significant difference in control losses can be attributed to the poor performance of the 15 × 2 (EOD) schedule. This schedule group had a lower maximum lambda and higher control losses (by approximately 0.5 units and 1.0 unit, respectively) than the other three schedule groups at the end of practice as well as at the start of mission testing.

CRC PC - The significant difference was due to the poorer performance of the 5 on/2 off schedule group which was almost 5% lower than performance of the other groups.

MAN PC - Switching-Manikin Task percent correct exhibited a marginally significant difference due to practice schedule. On the average, percent correct for subjects on the 7 × 2 per day schedule was approximately 1% higher.

Table 3. ANOVA Results Comparing End of Practice with First Mission by Testing Lapse.

		Schedule (S)		Lapse (L)		S x L		Trial (T)		S x T		L x T		S x L x T	
Task	Measure	F(3,88)	p	F(1,88)	p	F(3,88)	p	F(1,88)	p	F(3,88)	p	F(1,88)	p	F(3,88)	p
	LM	2.61	0.06	3.49	0.06	1.65	0.18	1.22	0.27	1.01	0.39	0.49	0.48	1.62	0.19
TRK	CL	3.40	0.02	2.36	0.13	0.93	0.43	0.00	0.99	0.94	0.42	0.13	0.72	0.68	0.57
	RM	1.03	0.38	0.04	0.84	0.90	0.45	0.05	0.82	1.16	0.33	0.09	0.77	0.12	0.95
	RT	1.92	0.13	0.06	0.81	0.23	0.87	0.58	0.45	0.83	0.48	1.24	0.27	0.79	0.50
MTX	PC	0.52	0.67	0.19	0.66	0.19	0.90	6.61	0.01	0.77	0.51	0.41	0.53	0.21	0.89
	TP	1.36	0.26	0.08	0.78	0.03	0.99	1.22	0.27	1.20	0.32	0.02	0.90	0.73	0.54
	RT	0.62	0.61	0.20	0.66	0.48	0.70	6.95	0.01	2.91	0.04	0.07	0.79	0.91	0.44
STN	PC	1.05	0.37	0.71	0.40	1.70	0.17	1.24	0.27	1.04	0.38	1.82	0.18	4.97	0.00
	TP	0.72	0.54	0.23	0.63	0.34	0.80	15.15	0.00	4.37	0.01	0.14	0.71	1.70	0.17
	RT	0.66	0.58	0.03	0.86	0.53	0.66	7.70	0.01	0.42	0.74	3.96	0.05	2.02	0.12
CRC	PC	6.09	0.00	6.36	0.01	4.30	0.01	0.04	0.84	0.26	0.85	0.12	0.73	0.71	0.55
	TP	0.45	0.72	0.10	0.75	1.01	0.39	5.89	0.00	0.83	0.48	0.92	0.34	3.02	0.03
MAN	RT	0.14	0.93	0.49	0.48	0.37	0.77	1.82	0.18	0.41	0.75	0.02	0.89	1.48	0.23
	PC	2.78	0.05	1.18	0.28	0.90	0.45	0.54	0.46	0.20	0.90	0.53	0.47	1.06	0.37
MTH	RT	0.70	0.55	1.39	0.24	2.68	0.05	5.22	0.02	1.91	0.13	0.65	0.42	1.43	0.24
	PC	0.88	0.45	0.63	0.43	0.06	0.98	3.07	0.08	1.87	0.14	0.01	0.91	0.78	0.51
	CL	0.38	0.77	1.44	0.23	0.12	0.95	0.57	0.45	1.31	0.28	3.97	0.05	0.50	0.68
	RM	1.19	0.32	0.07	0.80	0.31	0.82	3.89	0.05	1.63	0.19	0.90	0.35	0.20	0.90
DUL	RT	0.38	0.77	0.40	0.53	0.48	0.70	3.54	0.06	0.41	0.75	0.76	0.39	0.07	0.97
	PC	1.17	0.33	1.04	0.31	0.12	0.95	0.52	0.47	1.29	0.28	0.23	0.63	3.17	0.03
	TP	0.33	0.81	0.14	0.71	0.62	0.60	5.06	0.03	0.79	0.50	1.80	0.18	0.15	0.93
FAT	RP	2.34	0.08	0.38	0.54	0.66	0.58	5.16	0.03	1.54	0.21	0.01	0.91	0.66	0.58

Analysis of Testing Lapse from the end of practice to the start of mission revealed two significant differences in overall performance between the 3-day lapse and the 5-day lapse. Appendix G provides charts which clarify these differences if one visually averages data from the end of practice and MD-1 for each testing lapse.

TRK LM - A marginally significant difference was found for maximum lambda; subjects in the 5-day lapse group averaged slightly higher than those in the 3-day lapse group.

CRC PC - Continuous recognition percent correct was higher for subjects in the 3-day lapse group.

A significant Schedule × Lapse interaction indicates that a specific group (or groups) comprising a particular combination of practice schedule and testing lapse was different from other groups. There are eight unique combinations (omitting Group E) of these factors with overlap occurring between Groups A1 and A2, between Groups F and J, and between Groups G and K. Two measures yielded a significant Schedule × Lapse interaction. The data in Appendix E may help clarify these differences.

CRC PC - The significant difference in Continuous Recognition percent correct is due to the significantly poorer performance of Group D for the datapoints under consideration.

MTH RT - The differences in Mathematical Processing mean RT are marginally significant. Any strong effect is washed out by the large spread of the data across the groups.

The Trial factor tests whether there was an overall change in performance from end of practice to Mission Trial 1. These effects are best observed in the charts in Appendix F or G as the overall difference between end of practice and the first mission trial. Nine measures yielded significant differences. In general, both testing lapses created a negative impact on performance.

MTX PC - An overall performance decrease was seen across the testing lapse.

STN RT, TP - Sternberg response time exhibited a significant increase while throughput showed a corresponding decrease.

CRC RT, TP - Although Continuous Recognition response time was significant, the difference from the end of practice to the start of mission testing had little practical significance. Likewise, the statistically significant increase in throughput was negligible.

MTH RT - Mathematical Processing response time increased significantly following the lapse.

DUL RM, TP - The difference in the Dual-Tracking RMS error was marginally significant and cannot be observed in the charts. Dual-Memory Search demonstrated a slight, yet significant, increase in throughput.

FAT RP - A significant increase in Fatigue scores from the end of practice to the start of mission testing tends to imply that subjects grew more tired of the experiment following the testing lapse.

The Schedule × Trial interaction provides information about the relative impact of schedule differences across testing lapses. A significant difference for this effect would suggest that the practice schedules generated different levels of learning retention. Appendix F provides charts depicting this interaction. Only two performance measures were significant, indicating that, for the various schedules, there were minimal differences in the patterns of performance retention across the lapse.

STN RT, TP - Subjects on the 15 × 2 practice schedule maintained or improved Sternberg response time in comparison with the other schedule groups where response time increased. Sternberg throughput reflected these differences in RT; the 15 × 2 practice schedule showed slight improvement across the lapse.

The important test with respect to the differential impact of the 3-day vs. 5-day lapse is the Lapse × Trial interaction. Again, the charts in Appendix G illustrate this effect. Only two performance measures were statistically significant.

CRC RT - From a practical standpoint, there was no difference due to the 3-day vs. 5-day testing lapse.

DUL CL - Control losses for the group of subjects with a 3-day lapse decreased while those for the group with a 5-day lapse increased.

The final interaction examined was Schedule × Lapse × Trial. A significant interaction indicates a combined differential impact of practice schedule and testing lapse across the two trials. Recall that there are eight unique combinations of Schedule and Testing Lapse. The specific differential impact of the eight unique combinations can be determined through a tedious examination of the summary data in Appendix E. Three performance measures yielded significant differences.

STN PC - The significant differences can be attributed to the patterns in performance of Groups C and D. At the end of practice, all groups were tightly clustered. At the start of mission testing, Groups C and D exhibited a noticeable decline in performance while all other groups remained stable or improved slightly.

CRC TP - Although statistically significant, the differences among the eight group combinations appeared fairly consistent across trials. Not only did the spread of the data seem fairly constant, but the relative ordering of the practice schedules within each trial differed little.

DUL PC - The only noticeable difference was the lower performance of the combined Groups of A1 and A2 (primarily the poor performance of Group A2) at the end of practice. At the start of mission testing, no differences were apparent among the groups.

4.6 Mission Schedules

Two schedules were examined for mission testing: one trial every day for 5 days (ED Mission) and two trials every other day across 5 days (EOD Mission). Appendix H presents charts summarizing the data in terms of mission schedule. Again, Group E was omitted and data were collapsed across all remaining groups. Visual examination of the graphs reveals that subjects on both schedules followed the same pattern of performance during mission testing for virtually every task. Differences that do exist appear to be primarily due to random sample group differences. Only one performance measure yielded patterns worth noting.

DUL CL - Performance differences at the start of mission testing reflected similar differences evident at the end of practice. The ED Mission subjects had higher control losses than the EOD Mission subjects. However, throughout the mission trials, performance of the ED Mission subjects improved and control losses decreased such that performance was virtually identical to the performance of the EOD Mission subjects by Mission Day - 3.

4.7 Stability and Reliability

Intertrial correlations and Lord and Novick (1968) reliabilities were computed for the five mission testing sessions (Table 4). Lord and Novick (1968) reliabilities, computed as the ratio of (Between-subject Variability minus Within-subject Variability) divided by Between-subject Variability, had similar magnitude and paralleled the intertrial correlations.

Table 4. Intertrial Correlations and Lord & Novick Reliabilities.

Task -- Measure	Mission Days	
	Correlation	Reliability
Tracking -- Lambda (max)	0.74	0.73
Tracking -- Mean Lambda	0.78	0.77
Tracking -- Control Losses	0.66	0.66
Tracking -- RMS Error	0.73	0.72
Matrix -- Mean RT	0.89	0.89
Matrix -- Percent Correct	0.65	0.62
Matrix -- Throughput	0.92	0.91
Memory -- Mean RT	0.77	0.73
Memory -- Percent Correct	0.39	0.38
Memory -- Throughput	0.79	0.74
Con. Rec. -- Mean RT	0.92	0.91
Con. Rec. -- Percent Correct	0.73	0.72
Con. Rec. -- Throughput	0.91	0.90
Manikin -- Mean RT	0.92	0.90
Manikin -- Percent Correct	0.62	0.60
Manikin -- Transition RT	0.89	0.88
Manikin -- Transition PC	0.50	0.50
Math -- Mean RT	0.91	0.89
Math -- Percent Correct	0.59	0.52
Math -- Transition RT	0.87	0.84
Math -- Transition PC	0.47	0.42
Dual -- Control Losses	0.70	0.70
Dual -- RMS Error	0.83	0.80
Dual MS -- Mean RT	0.84	0.84
Dual MS -- Percent Correct	0.42	0.36
Dual MS -- Throughput	0.85	0.84
Fatigue -- Response	0.47	0.59
Fatigue -- RT	0.44	0.47

The very good to excellent levels of differential stability and reliability (typically on the order of 0.80 or above), obtained for at least one measure on all of the performance tests, justifies each test's inclusion in the NASA Performance Assessment Battery.

4.8 Performance Database

Analyses of the various performance measures revealed few significant differences due to the various practice schedules. With the omission of Group E data from the analysis, the number of significant differences decreased. Therefore, to construct the performance database for classifying astronaut performance, data were collapsed across 96 subjects combining all practice schedule groups except Group E.

Overall performance measures for the collapsed database using Training, Start and End of Practice, and the five Mission days are illustrated in Figures 6 through 21. The trends evident in the graphs of the collapsed data mirror those evident in the graphs depicting the various schedules (Appendixes F through H). Generally, little difference is seen from training to the start of practice, while overall performance improved from the start to the end of practice. Performance across mission days (MD-1 through MD-5) remained highly stable for almost all measures.

Matrix throughput and Continuous Recognition throughput showed slight continued improvement throughout mission testing as did the response times for the Switching tasks. Improvement in Dual Task control losses was also observed during the early mission trials.

Appendix I summarizes the performance database using box-and-whisker plots. Within these plots, each trial is represented by several statistics displayed as a box with vertical lines, or "whiskers", extending from either end. The bottom and top edges of the box represent the 25th and 75th percentiles (i.e., first and third quartiles), respectively; the median is drawn as a center horizontal line in the box. The whiskers extend from the edges of the box to the minimum and maximum data points provided these do not exceed 1.5 times the interquartile range (i.e., the distance between the first and third quartiles). Extreme values within 3 interquartile ranges are shown as "*". For more extreme values, "o" is used. Thus, the plots provide a clear visual image of the central tendency and spread for the key performance measures during training, practice and mission testing.

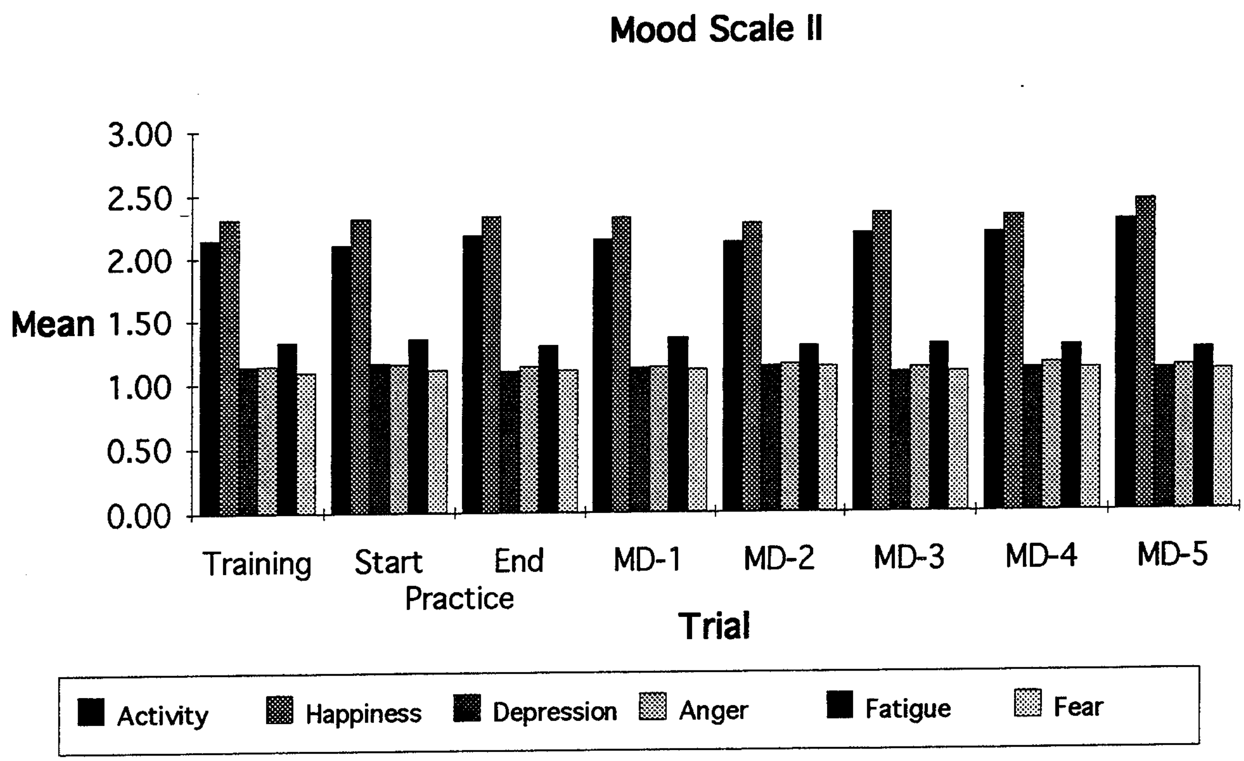


Figure 6. Response by Trial for Mood Scale II.

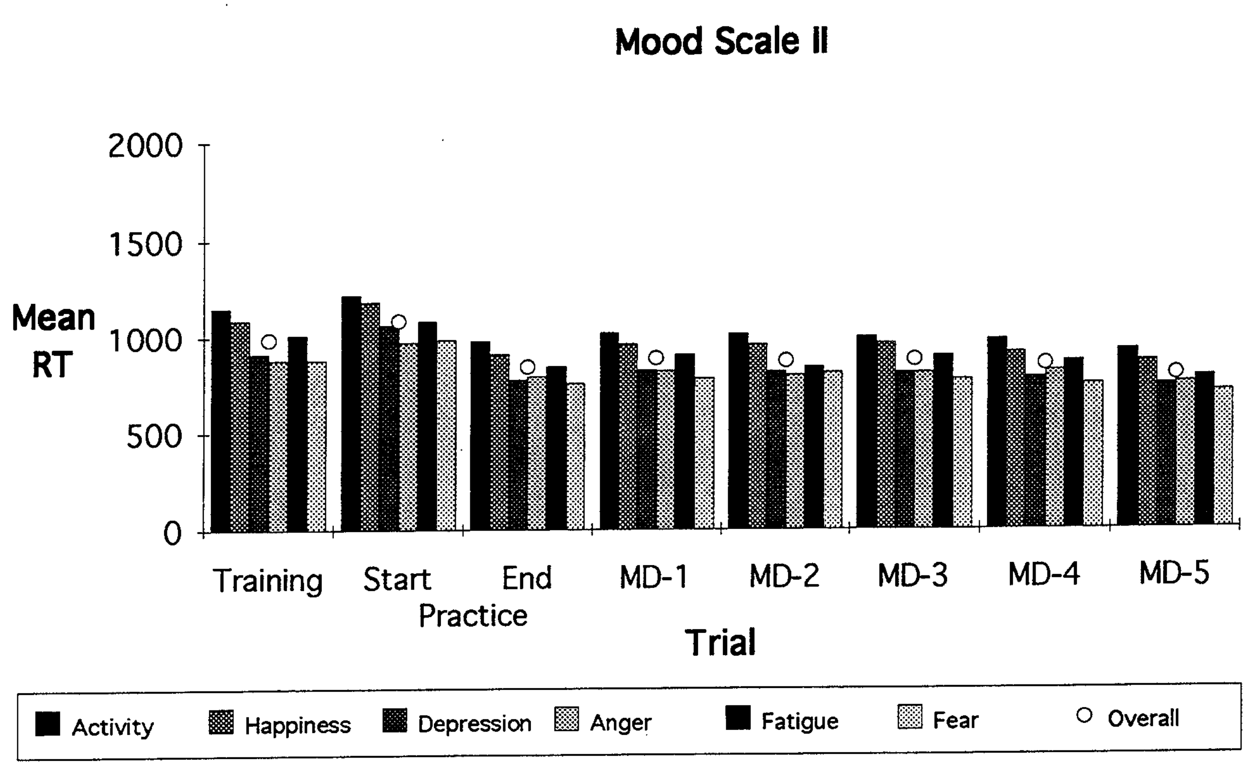


Figure 7. Mean Response Time by Trial for Mood Scale II.

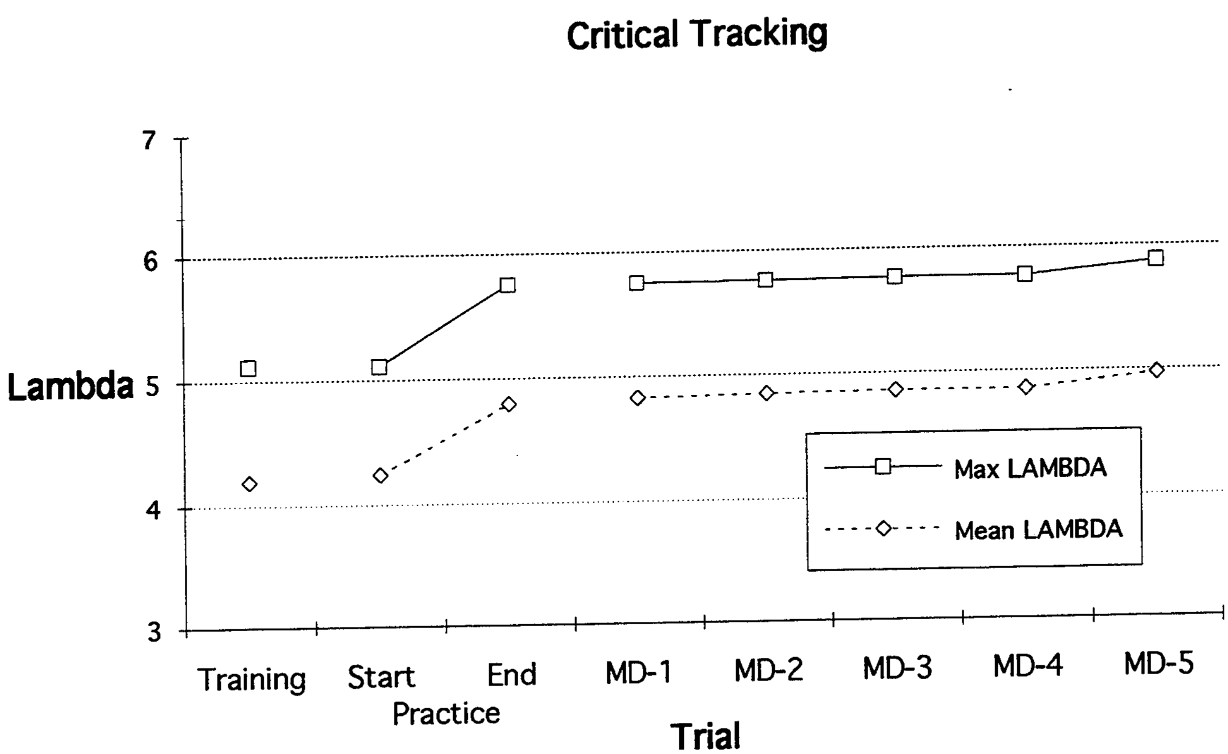


Figure 8. Maximum Lambda and Mean Lambda by Trial for Critical Tracking.

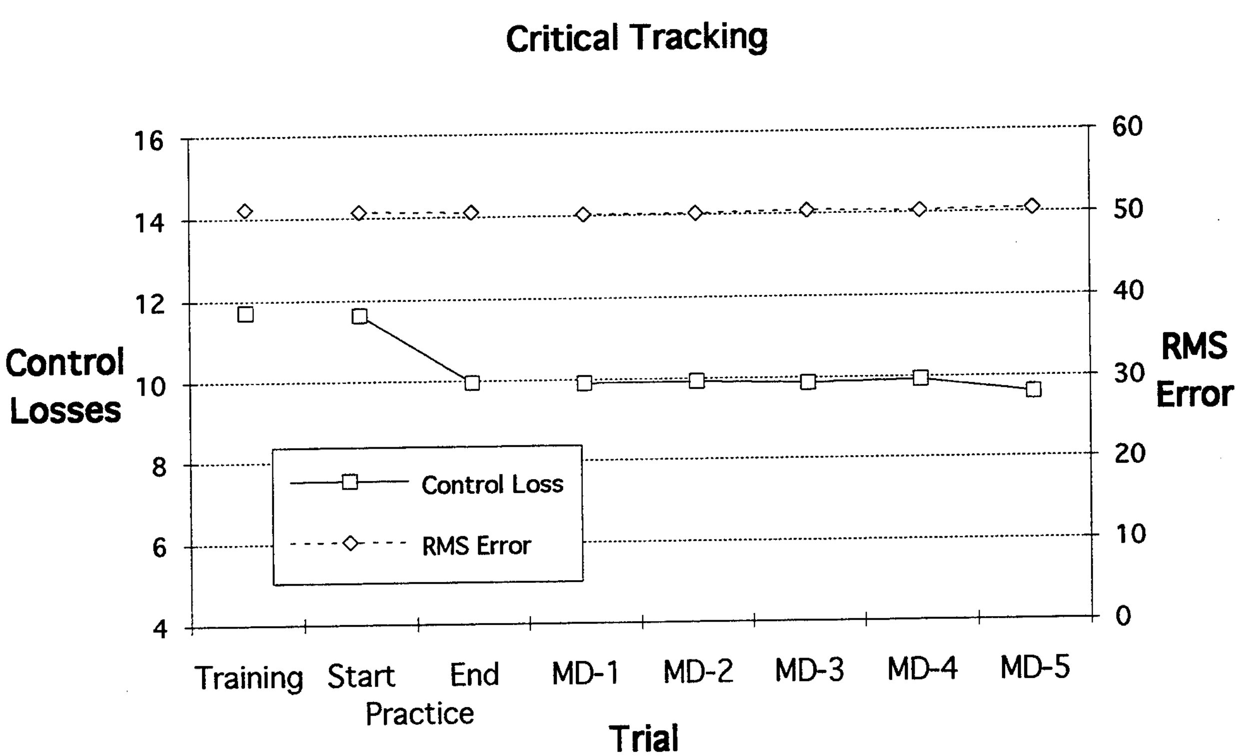


Figure 9. Control Losses and RMS Error by Trial for Critical Tracking.

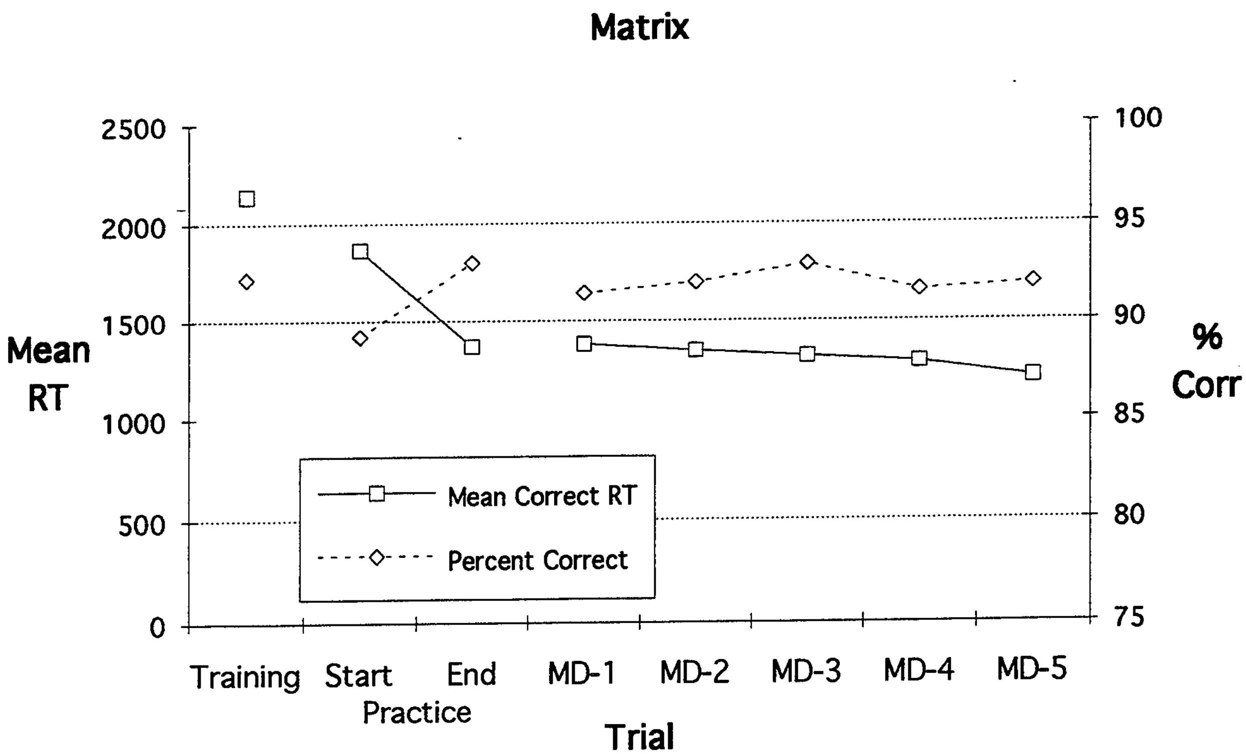


Figure 10. Mean Response Time and Percent Correct by Trial for Matrix.

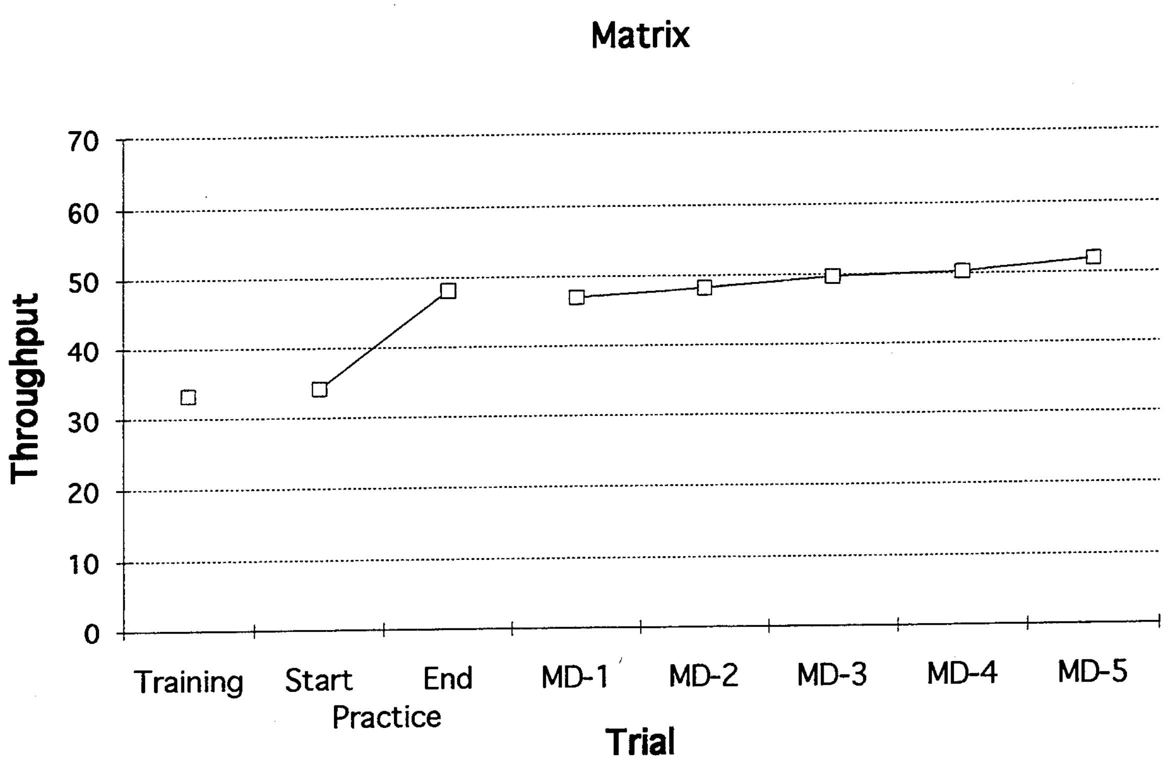


Figure 11. Throughput by Trial for Matrix.

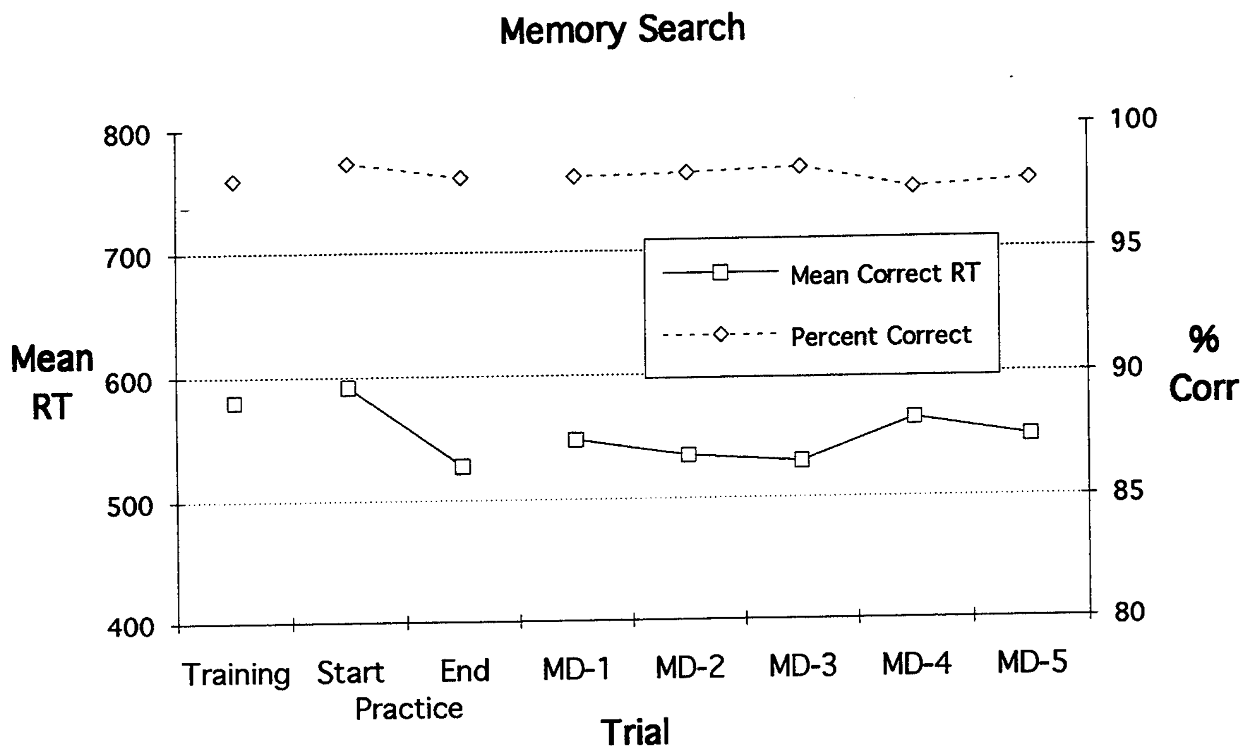


Figure 12. Mean Response Time and Percent Correct by Trial for Memory Search.

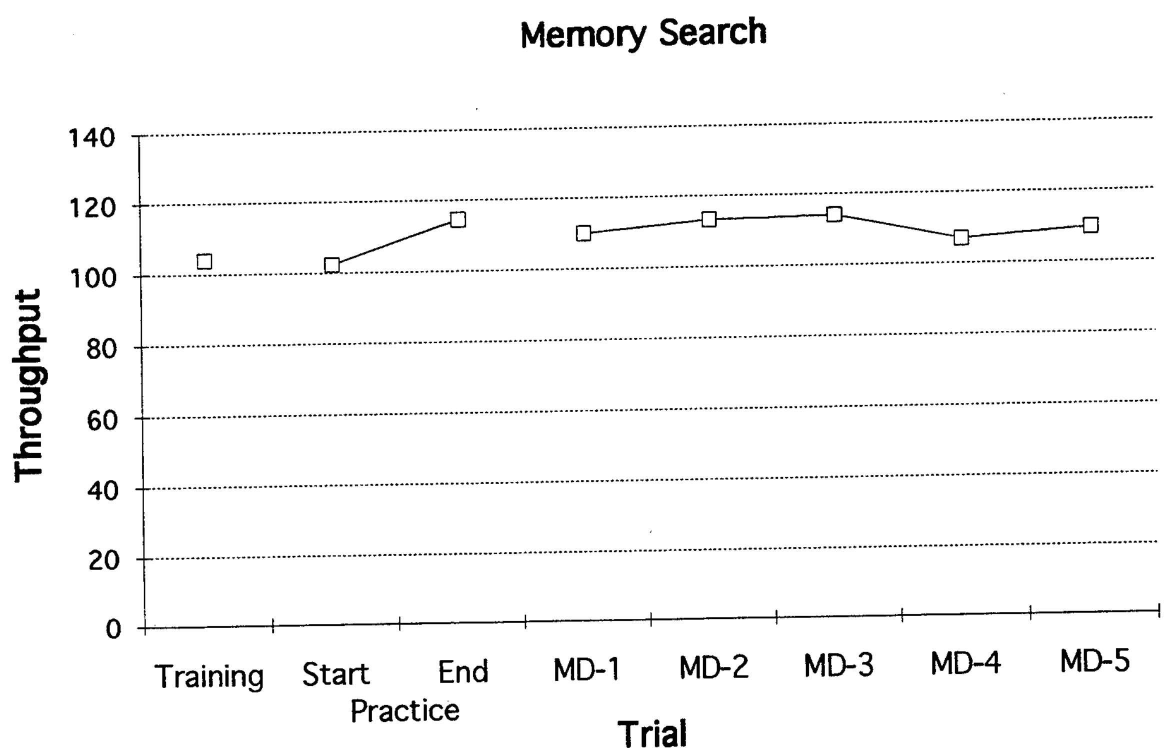


Figure 13. Throughput by Trial for Memory Search.

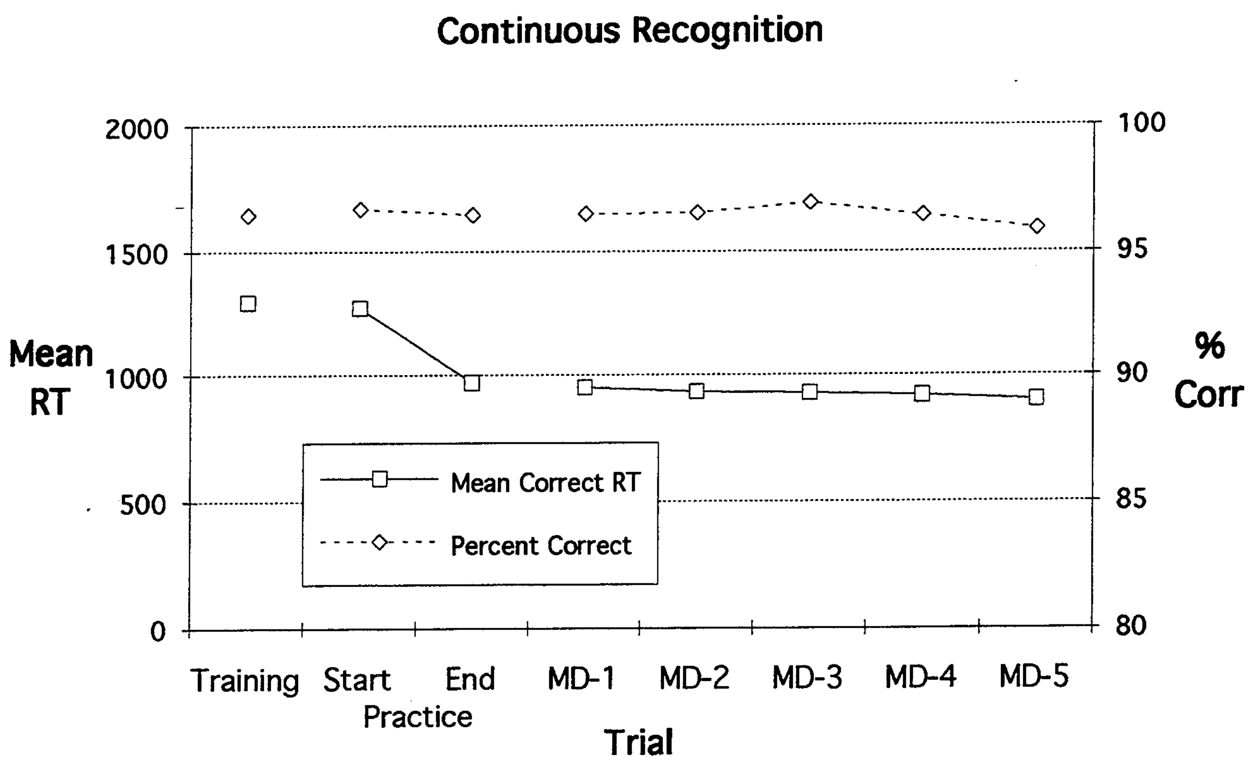


Figure 14. Mean Response Time and Percent Correct by Trial for Continuous Recognition.

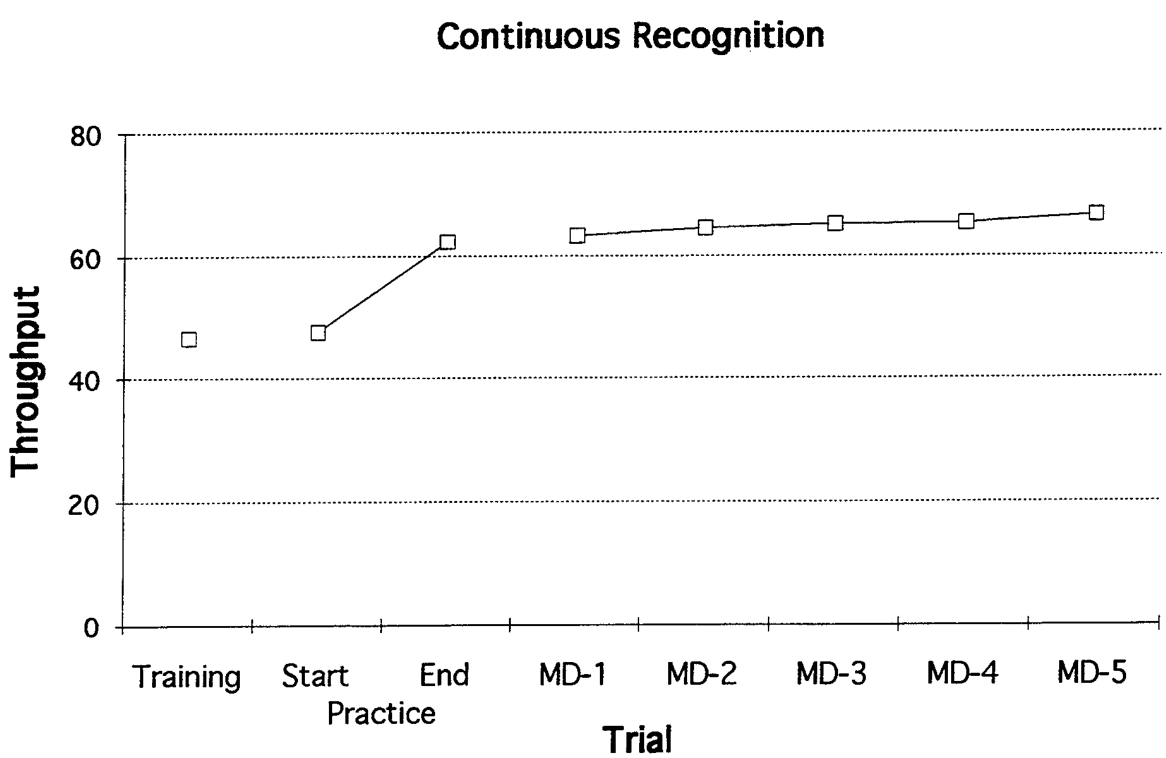


Figure 15. Throughput by Trial for Continuous Recognition.

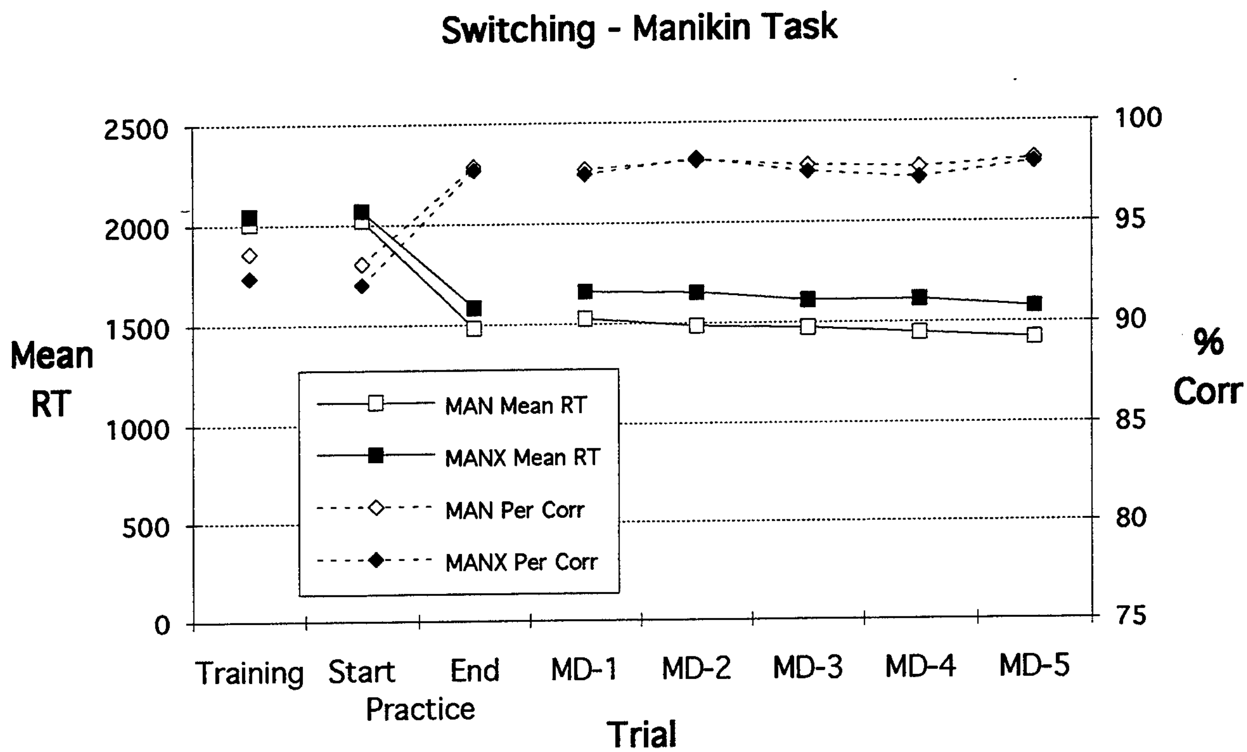


Figure 16. Mean Response Time and Percent Correct by Trial for Switching - Manikin Task.

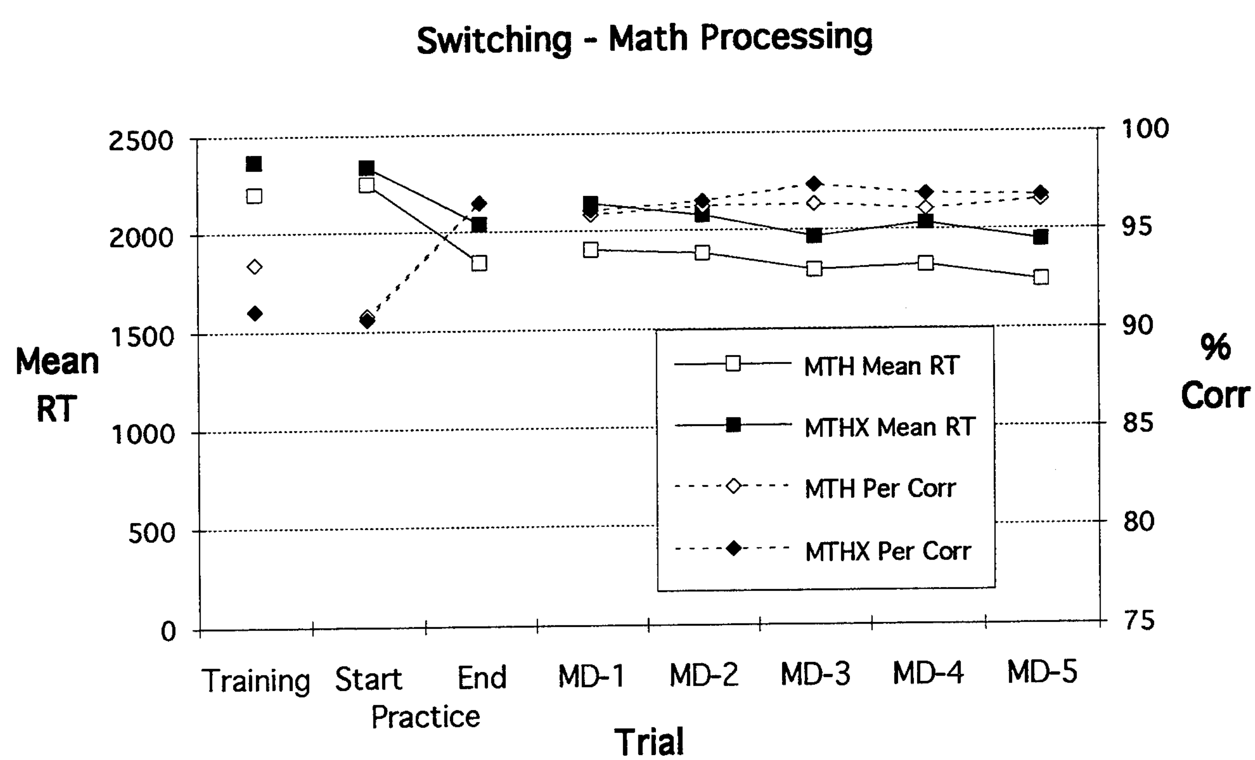


Figure 17. Mean Response Time and Percent Correct by Trial for Switching - Math Processing.

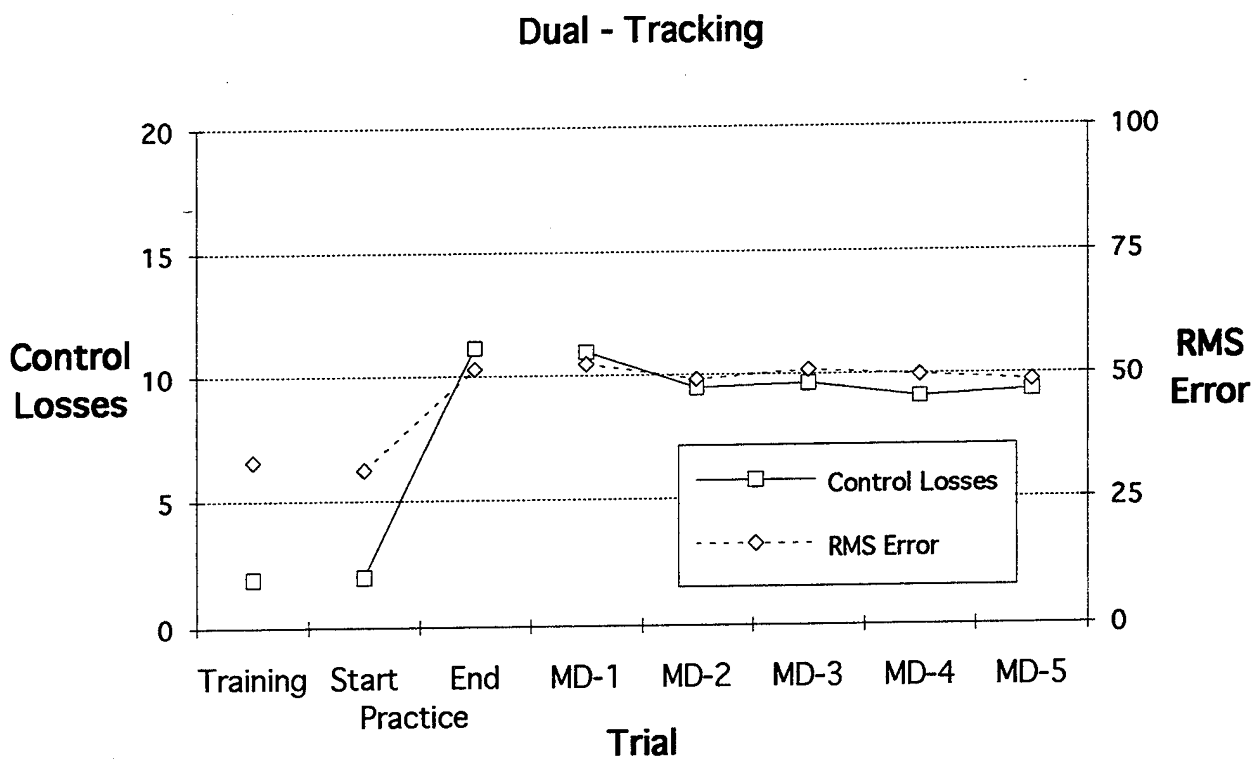


Figure 18. Control Losses and RMS Error by Trial for Dual - Tracking.

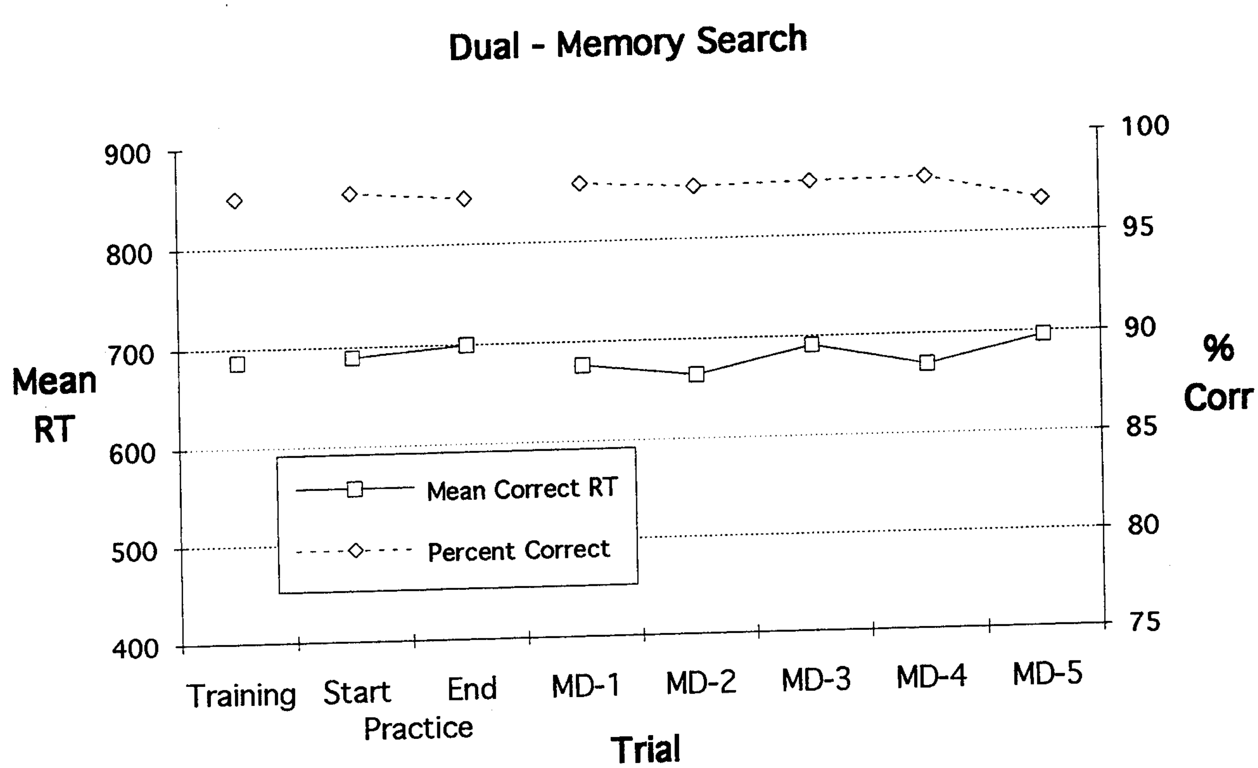


Figure 19. Mean Response Time and Percent Correct by Trial for Dual - Memory Search.

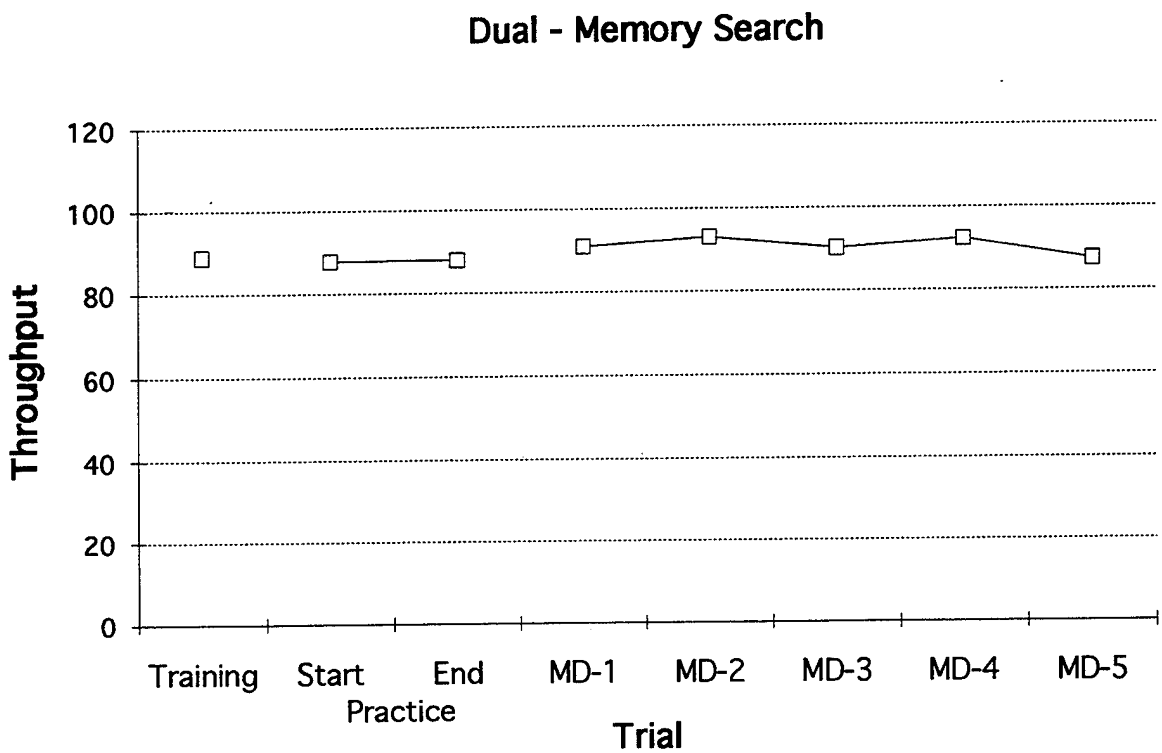


Figure 20. Throughput by Trial for Dual - Memory Search.

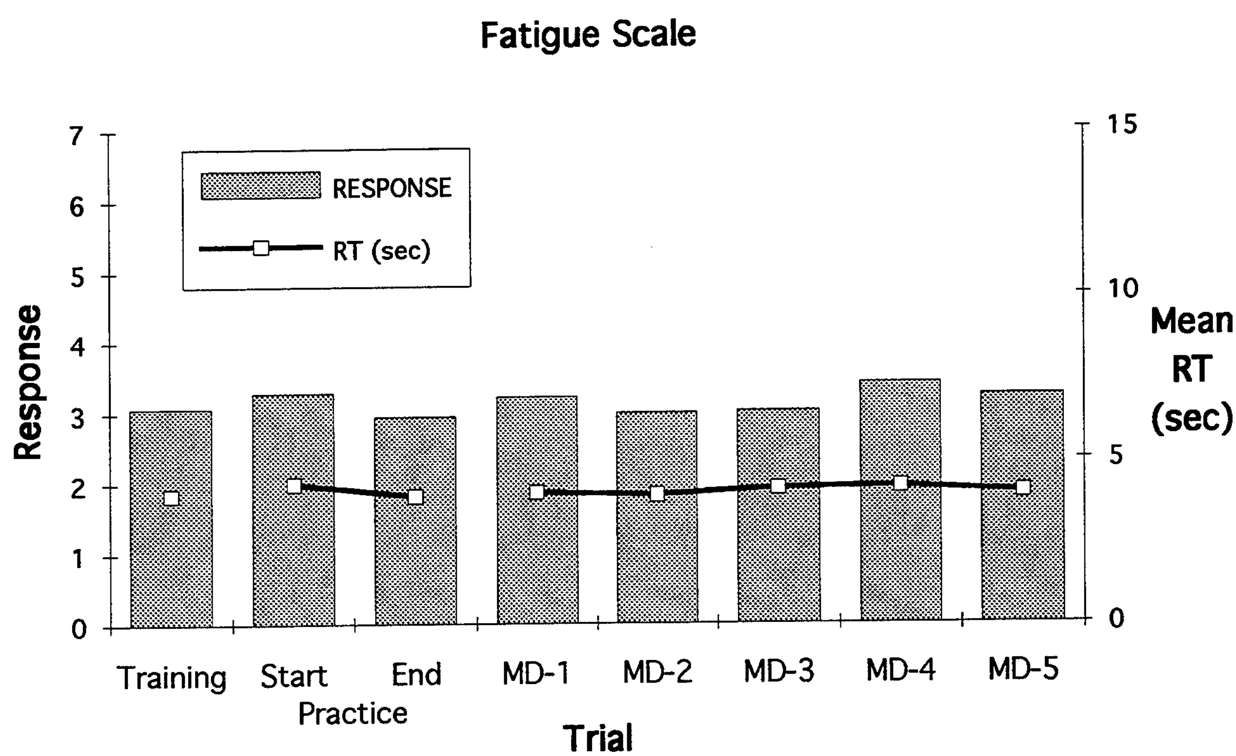


Figure 21. Response and Mean Response Time by Trial for Fatigue Scale.

To further summarize the database, quantiles were computed at 20% intervals (Table 5). Data were arranged such that the 0%ile value represents the poorest performance obtained for that trial while the 100%ile represents the best performance. In other words, 0%ile RTs would be the largest values or the slowest RTs, but 0%ile PCs would be the smallest values representing the lowest accuracy. The table of quantiles presents discrete values which can be used to classify performance within ranges as follows:

0% - 20%: Very Poor Performance
20% - 40%: Poor Performance
40% - 60%: Average Performance
60% - 80%: Good Performance
80% - 100%: Very Good Performance.

Table 5. Performance Percentiles - All Subjects.

		Train	Start	End	MD1	MD2	MD3	MD4	MD5
TRK LM	0%	3.9	4.1	4.4	4.4	3.7	4.4	3.6	4.7
	20%	4.7	4.7	5.2	5.3	5.3	5.2	5.2	5.3
	40%	5.0	5.0	5.5	5.6	5.6	5.6	5.5	5.6
	60%	5.3	5.2	5.9	5.9	6.0	5.9	5.9	6.1
	80%	5.6	5.6	6.3	6.3	6.2	6.4	6.3	6.4
	100%	6.5	6.4	7.7	7.5	7.5	7.1	7.4	7.5
TRK ML	0%	3.1	3.2	3.7	3.6	2.8	3.7	2.7	3.5
	20%	3.8	3.8	4.3	4.3	4.4	4.4	4.3	4.4
	40%	4.1	4.0	4.6	4.6	4.8	4.6	4.7	4.7
	60%	4.3	4.3	5.0	5.0	5.0	4.9	5.0	5.1
	80%	4.6	4.8	5.3	5.3	5.4	5.4	5.4	5.7
	100%	5.4	5.2	6.1	6.5	6.1	6.4	6.5	6.6
TRK CL	0%	17	15	14	14	18	13	19	13
	20%	13	13	11	11	11	11	11	11
	40%	12	12	11	10	10	10	10	10
	60%	11	11	10	9	9	9	9	9
	80%	11	10	9	9	9	9	9	8
	100%	9	9	7	7	7	7	7	7
TRK RM	0%	64	71	74	79	81	85	74	79
	20%	58	59	58	56	57	58	58	57
	40%	54	52	53	52	52	52	51	51
	60%	49	48	48	47	47	47	48	47
	80%	46	43	44	44	43	42	42	44
	100%	31	30	30	32	29	30	28	30

		Train	Start	End	MD1	MD2	MD3	MD4	MD5
MTX RT	0%	7804	6019	6174	5683	5603	4832	6852	6221
	20%	2912	2374	1642	1684	1746	1727	1648	1439
	40%	1988	1777	1282	1263	1252	1228	1192	1118
	60%	1445	1360	1036	1058	1037	1038	1001	964
	80%	1175	1146	889	879	884	852	842	813
	100%	798	763	644	625	607	563	515	565
MTX PC	0%	72	71	73	70	65	79	64	73
	20%	88	82	89	88	88	88	87	87
	40%	92	89	93	91	91	93	92	92
	60%	94	93	95	94	95	96	95	95
	80%	97	95	97	97	97	98	98	97
	100%	100	100	100	100	100	100	100	100
MTX TP	0%	7	10	10	10	11	12	9	10
	20%	19	22	32	32	31	33	33	36
	40%	27	30	42	42	43	42	42	47
	60%	39	38	52	50	53	54	53	56
	80%	46	46	64	60	63	64	65	66
	100%	64	71	85	88	92	92	100	94

Table 5. Performance Percentiles - All Subjects (cont.).

		Train	Start	End	MD1	MD2	MD3	MD4	MD5
	0%	822	901	874	1086	1070	761	957	977
	20%	649	657	578	593	585	588	608	610
STN RT	40%	579	597	534	552	520	534	550	540
	60%	542	554	502	514	497	498	526	511
	80%	504	527	455	475	471	464	486	480
	100%	413	401	369	384	380	386	372	406
	0%	85	91	90	87	90	91	88	87
	20%	97	97	97	97	96	97	95	96
STN PC	40%	98	99	98	99	99	99	97	97
	60%	99	100	99	99	99	99	99	99
	80%	99	100	99	100	100	100	100	100
	100%	100	100	100	100	100	100	100	100
	0%	71	64	68	55	52	75	61	61
	20%	92	91	103	99	102	101	95	96
STN TP	40%	100	100	111	108	114	111	107	109
	60%	108	107	118	115	118	120	113	116
	80%	117	113	129	123	126	127	121	122
	100%	138	141	148	141	150	150	155	146

	0%	2166	2130	1861	1616	1460	1622	1814	1729
	20%	1537	1581	1169	1137	1149	1125	1114	1108
CRC RT	40%	1337	1284	1016	987	963	952	942	924
	60%	1192	1154	905	877	855	839	834	831
	80%	1023	1011	748	769	744	742	747	718
	100%	676	657	576	550	548	541	528	530
	0%	72	74	77	77	76	67	78	69
	20%	95	96	94	94	95	95	95	94
CRC PC	40%	97	97	97	97	97	97	97	97
	60%	98	99	99	99	99	99	99	99
	80%	99	100	100	99	100	100	100	99
	100%	100	100	100	100	100	100	100	100
	0%	27	28	32	36	37	35	32	32
	20%	36	36	49	50	49	52	53	53
CRC TP	40%	42	44	58	59	57	60	61	60
	60%	49	50	65	65	68	67	68	70
	80%	56	57	73	77	77	79	76	80
	100%	88	91	103	104	109	108	110	103

Table 5. Performance Percentiles - All Subjects (cont.).

		Train	Start	End	MD1	MD2	MD3	MD4	MD5
MAN RT	0%	3077	3358	2330	2585	2570	2611	2559	2382
	20%	2418	2398	1765	1870	1861	1746	1733	1736
	40%	2033	2093	1580	1672	1549	1567	1535	1526
	60%	1803	1862	1346	1375	1328	1357	1308	1285
	80%	1622	1626	1152	1184	1131	1162	1131	1089
	100%	1111	1055	929	872	752	740	772	765
MAN PC	0%	65	66	84	83	74	87	79	83
	20%	90	88	97	96	97	96	96	97
	40%	96	95	99	99	98	98	98	98
	60%	98	98	99	100	100	100	99	100
	80%	99	100	99	100	100	100	100	100
	100%	100	100	100	100	100	100	100	100
MAN RTX	0%	3062	3569	2743	2763	2923	2953	2933	2904
	20%	2429	2482	1880	2022	2070	1937	2042	1950
	40%	1961	2173	1634	1748	1699	1626	1627	1648
	60%	1822	1842	1397	1448	1412	1434	1395	1339
	80%	1686	1675	1260	1256	1236	1251	1256	1208
	100%	1280	1280	957	982	930	893	993	906
MAN PCX	0%	58	57	81	74	69	79	78	82
	20%	86	85	97	96	97	96	94	96
	40%	94	95	98	100	100	100	99	100
	60%	98	100	100	100	100	100	100	100
	80%	100	100	100	100	100	100	100	100
	100%	100	100	100	100	100	100	100	100
MTH RT	0%	3144	3386	2745	2875	2792	2914	2863	2809
	20%	2644	2677	2254	2270	2234	2156	2148	2042
	40%	2228	2369	1915	1965	2001	1846	1935	1838
	60%	2069	2122	1702	1790	1735	1657	1679	1570
	80%	1864	1858	1493	1513	1520	1433	1444	1412
	100%	1389	1010	1049	1049	1049	1032	1035	1054
MTH PC	0%	72	53	80	81	72	83	73	72
	20%	90	86	96	93	94	94	94	94
	40%	94	92	97	96	97	96	96	97
	60%	96	96	98	98	99	99	98	99
	80%	98	98	99	100	100	100	100	100
	100%	100	100	100	100	100	100	100	100
MTH RTX	0%	3637	3655	3427	3456	3408	3122	3270	3032
	20%	2907	2746	2488	2474	2487	2335	2439	2284
	40%	2413	2435	2056	2219	2150	1986	2113	2005
	60%	2151	2195	1910	2068	1954	1824	1819	1803
	80%	1960	1851	1693	1708	1668	1575	1624	1553
	100%	1519	1029	1050	1184	1061	1061	1030	1112
MTH PCX	0%	58	52	81	79	67	88	75	57
	20%	86	82	95	94	94	95	94	95
	40%	92	92	97	97	97	97	97	97
	60%	95	96	98	100	100	100	100	100
	80%	98	100	100	100	100	100	100	100
	100%	100	100	100	100	100	100	100	100

Table 5. Performance Percentiles - All Subjects (cont.).

		Train	Start	End	MD1	MD2	MD3	MD4	MD5
DUL RT	0%	1178	1096	1474	1342	1312	1766	1424	1594
	20%	794	791	793	781	753	806	753	765
	40%	685	706	690	667	666	682	678	708
	60%	618	626	641	620	597	627	603	635
	80%	579	578	564	560	551	543	553	569
	100%	449	463	435	413	406	402	402	424
DUL PC	0%	76	88	89	90	83	89	90	84
	20%	96	97	96	96	96	96	96	95
	40%	98	98	97	98	98	97	98	97
	60%	99	99	98	99	99	99	99	98
	80%	99	100	99	100	100	100	100	99
	100%	100	100	100	100	100	100	100	100
DUL TP	0%	43	53	40	42	43	30	38	33
	20%	75	73	74	76	78	73	78	75
	40%	85	83	84	89	87	88	87	83
	60%	95	93	92	95	99	95	97	92
	80%	102	102	105	105	107	106	107	104
	100%	133	128	132	144	140	144	141	132
DUL CL	0%	38	42	59	51	81	60	54	54
	20%	3	3	21	18	14	15	14	15
	40%	1	1	11	11	8	9	8	7
	60%	0	0	6	6	4	5	5	5
	80%	0	0	3	2	1	2	1	1
	100%	0	0	0	0	0	0	0	0
DUL RM	0%	73	73	81	94	91	96	85	87
	20%	51	52	63	63	61	63	63	62
	40%	35	33	58	58	54	58	56	55
	60%	26	22	50	52	47	51	49	48
	80%	15	13	40	40	36	38	37	34
	100%	7	6	18	18	4	5	9	6
FAT RP	0%	7	7	7	7	7	7	7	7
	20%	5	4	4	5	4	4	4	4
	40%	4	3	4	3	3	3	3	3
	60%	3	3	3	3	3	3	3	2
	80%	2	2	2	2	2	2	2	2
	100%	1	1	1	1	1	1	1	1
FAT RT	0%	10.5	12.4	8.3	9.1	7.4	12.3	12.8	7.1
	20%	5.8	6.3	4.0	4.9	3.2	4.2	3.2	3.6
	40%	4.1	5.0	2.7	3.3	2.0	2.9	1.9	2.9
	60%	3.0	4.4	2.1	2.1	1.3	1.9	1.2	2.2
	80%	2.7	3.5	1.3	1.3	0.8	1.2	0.6	0.8
	100%	0.5	1.4	0.4	0.4	0.4	0.3	0.1	0.1

4.9 Subject Debriefing

Upon completing the experiment, subjects at the University of Oklahoma were asked to complete a two-page debriefing questionnaire. The questionnaire and a detailed summary of the results are presented in Appendix J. In total, 79 subjects completed the questionnaire. Responses for various questions do not always total 79 because some subjects provided more than one response for some questions.

Question 1:

Seventy-one subjects (90%) expressed a positive opinion about their participation in the experiment. Various subjects described their experience as challenging, interesting, enjoyable, etc. Thus, it is not surprising that all of the subjects replied that they would either redo the experiment or recommend it to a friend. However, 10 subjects attached restrictions to their "YES" response, such as different tasks or a shorter time frame. Fourteen of the 79 subjects (18%) had something negative to report about the experiment (boring, monotonous, tiresome or too long).

Question 2:

Regarding task preferences, the Sternberg Memory Search task was the most preferred task by 20 subjects followed by Continuous Recognition and Switching preferred by 18 participants each. Almost the same number of subjects (16) showed a preference for the Dual Task. The task listed as most preferred by the fewest number of subjects (10) was Matrix. The reason most frequently cited for selecting the Memory Search (15 responses) and Continuous Recognition (7 responses) tasks as most preferred was that these tasks were easy. At the same time, the somewhat complex nature of the Switching and Dual tasks seems to be the reason why other subjects prefer these tasks. Nine subjects who favored Switching and nine who favored the Dual Task described them as challenging.

Question 3:

Tracking was by far the least preferred task (28 responses). Fifteen subjects characterized the task as too hard. Five participants became frustrated when performing this task. The Dual Task was the second least preferred task for primarily the same reason. Ten of the sixteen negative responses for this task were due to its difficult nature. Negative opinions were distributed about equally for the rest of the tasks.

Almost all subjects (75) thought that they received enough instruction to perform the tasks. Four subjects thought that they were not instructed adequately for some of the tasks (primarily Switching). However, 32 participants departed from a task's standard instructions on at least one trial during their participation. Such departures usually involved using the wrong keys or the wrong fingers when responding. For eleven subjects, this digression occurred and was corrected during the familiarization period of the first week (training).

Question 4:

Most of the participants (69) reported that they thoughtfully and accurately responded on the MOOD and FATIGUE scales, although hitting a wrong key occasionally was a problem noted by 13 subjects. When subjects were asked to comment on these two scales, a variety of opinions was obtained. Most subject comments indicated that the two scales did not cover all possible mood and/or fatigue levels experienced during testing.

Question 5:

In terms of special strategies developed by the subjects during the course of the study, it is interesting to note that specific strategies were developed by some subjects for almost all tasks. For the matrix task, the two more popular strategies were viewing the overall matrix as a whole (18 responses) and looking at only 2 or 3 matrix cells (14 responses). These two strategies coupled with a third strategy of looking for patterns or designs within the matrix, indicated that at least these subjects did not memorize the exact locations of all highlighted cells in the matrix.

In the Sternberg task, the basic strategy mentioned by many subjects (28 responses) was a continuous mental rehearsal of the positive set throughout performance of the task. A second strategy (22 responses) was to generate a word from the letters of the positive set. For Continuous Recognition, simply concentrating on the bottom numbers of the ratios was mentioned by 16 respondents, whereas 16 others developed the rehearsal strategy of repeating aloud or in mind the displayed numbers. Imagining their body in the manikin's position was the strategy most often indicated by the subjects for the manikin portion of the switching task. Another fairly common strategy for the same task was to memorize the mapping of all stimuli to their responses (8 responses).

Questions 6 and 7:

When subjects were asked to evaluate the testing environment, 41 of them either made no comments or were satisfied. However, some comments were made concerning screen glare (13 subjects) and distraction by subjects entering and leaving the testing room.

5.0 CONCLUSIONS

This section addresses each of the specific research objectives stated in Section 2.0. In addition to conclusions derived from the data at this point in time, specific recommendations for further analysis of the data are offered. Finally, recommendations for courses of action based on the data are presented.

(1) Optimal Practice Schedule. Several different schedules were able to produce quite similar improvements in performance from the start to the end of practice. Therefore, in terms of group means, one cannot select a single practice schedule as optimal. When Group E (2 on/5 off, only 6 total trials) data were excluded from the analyses, only 3 of the 22 key performance measures yielded a significant practice schedule by trial interaction. Matrix throughput did not improve as rapidly during practice for the subjects on the 15-day, two trials every-other-day schedule. However, all subjects in this group received initial instructions to view the matrix patterns rapidly and to make quick responses. Thus, the average throughput for this group at the end of training was already quite high and did not increase as much during practice as did throughput for the other groups. The differences in Continuous Recognition throughput had little practical significance. Although not reflected in terms of any performance difference, subjects following the compressed schedule of two successive trials per day for 7 days registered higher Fatigue levels than did the other subjects.

(2) Effect of Testing Lapses. The testing lapse between the end of baseline practice and the start of mission testing resulted in a statistically significant performance decrement for 9 of the 22 measures analyzed. However, there was no appreciable differential impact of a 5-day lapse vs. a 3-day lapse with the exception of the number of control losses for the Dual Tracking task. In all cases, performance recovery was achieved by the second or third mission trial.

(3) Stability and Reliability of Test Measures. The very good to excellent levels of differential stability and reliability (0.78 to 0.92) obtained for at least one measure on all tasks justifies the inclusion of all selected tasks in the NASA PAWS.

(4) New Measure of Critical Tracking Performance. Computation of the Mean Lambda value achieved during a session provided a more stable measure of Critical Tracking performance. Correlation and reliability values improved from 0.69 for the previously used Maximum Lambda metric to 0.78 for the new measure.

(5) Reliability of the Testing Software. Excellent software reliability was confirmed with fewer than 35 missed tests out of 23,000 tests administered during 2,377 subject sessions (99.85% reliability). New or persistent problems were reported to the software developers and were remedied.

(6) Performance Database for Classifying Astronaut Performance. Performance data based on 101 female and male subjects ranging in age from 17 to 52 years were collected during training, baseline practice, and simulated mission testing periods. These data have been summarized in tabular and graphical form. Box-and-whisker plots for 24 key measures have been provided for the end of training, the start and end of practice, and five mission trials. Percentile cut points at the 0, 20, 40, 60, 80, and 100 percentile levels have also been provided for 28 measures at the same points in time. This database will serve as a valuable resource for comparisons of specific astronaut performance.

5.1 Research Recommendations

(1) Gender and Age Performance Variation. It is recommended that the impact of subject gender and age be determined through additional partitioning and analysis of the existing database.

(2) Stability and Reliability. Further analysis of the stability and reliability of the task measures as a function of the various practice schedules is warranted. One approach would be to compare stability and reliability at the end of training with the improved values at the end of practice and during mission testing for each practice schedule group.

(3) Multivariate Analyses. In order to explore the factor structure of the performance assessment battery in terms of the type of cognitive activity measured by each task, multivariate analyses such as cluster and/or factor analyses are recommended.

(4) Single-Subject Analyses. Due to the limited astronaut subject sample for IML-2, exploration of single-subject analyses of the data is recommended. This alternative may involve matching astronauts with specific individuals from the science support study who share the same general skill acquisition patterns and levels of performance and/or the same strengths and weaknesses on specific tasks.

(5) Task Software Evaluation. Continued evaluation and refinement of the task software should be pursued to confirm the already existing high reliability and to evaluate the most recent modifications.

5.2 Astronaut Testing Recommendations

(1) **Task Instructions.** The impact of varying the task instructions was observed during the Matrix task when some subjects were instructed to view the matrix patterns rapidly and to make quick responses. It highlights the importance of issuing identical instructions to the astronaut subjects as a single group. This procedure would allow individual questions to be answered uniformly for the benefit of the entire group. The instructions should be identical to those issued to the largest subject group from the support study in order to provide the largest amount of comparison data. The Switching task was the most problematic. The Manikin task was the primary source of the problem. All manikin views should be presented to each subject prior to the first session in order to confirm that the subject understands the concepts of the task.

(2) **Baseline Practice Schedules and Testing Lapses.** As long as an adequate number of baseline practice trials are obtained (minimum of 15), the practice schedule may be compressed (two trials per day) and/or interrupted for short periods (1 to 3 days). The impact of greater compression is not known. Both 3-day and 5-day testing lapses had a negative impact on mean performance for a number of tasks, although the impact did not appear greater for a 5-day lapse. This provides further support for the request that all testing lapses be minimized.

(3) **Mission Testing Schedules.** Performance data collected during two successive trials per day, every-other-day did not differ from data collected once per day, every day. Every-other-day testing thus appears justified. However, separating the two trials by a work shift (beginning and end of shift) may provide information on the fatiguing effects of work in space.

6.0 REFERENCES

Lord, F.M., and Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Schlegel, R.E., and Shehab, R.L. (1992). *Evaluation of astronaut practice schedules for the International Microgravity Laboratory (IML-2)*. Final Report, 1992 USAF-RDL Summer Faculty Research Program. Culver City, CA: Research and Development Laboratories.

APPENDIX A

SUBJECT CHARACTERISTICS

All Subjects

	Count	Mean Age	Std. Dev. Age	Min Age	Max Age
	101	26.0	8.1	17.0	52.0
A1	4	42.3	11.6	25.0	50.0
B	4	31.3	7.9	23.0	39.0
C	4	37.3	14.4	23.0	52.0
D	4	35.5	8.6	26.0	45.0
E	5	30.6	8.2	24.0	44.0
A2	20	24.2	6.1	19.0	44.0
F	10	23.9	6.1	19.0	39.0
G	10	20.8	3.0	17.0	26.0
H	9	23.0	3.8	18.0	30.0
I	10	24.0	5.1	18.0	32.0
J	10	25.0	3.9	20.0	30.0
K	11	24.1	7.3	19.0	44.0
Brooks	21	35.1	10.2	23.0	52.0
OU	80	23.7	5.3	17.0	44.0
15 ED	28	27.8	9.5	19.0	50.0
5 on/2 off	8	36.4	11.0	23.0	52.0
2 on/5 off	5	30.6	8.2	24.0	44.0
7 x 2 per day	41	23.5	5.5	17.0	44.0
15 x 2 (EOD)	19	23.5	4.4	18.0	32.0
3-day break	59	25.7	8.9	17.0	52.0
5-day break	42	26.6	6.9	18.0	45.0

Females

	Count	Mean Age	Std. Dev. Age	Min Age	Max Age
	36	23.4	5.3	17.0	39.0
A1	0			0.0	0.0
B	2	30.0	9.9	23.0	37.0
C	2	25.0	2.8	23.0	27.0
D	0			0.0	0.0
E	2	26.5	3.5	24.0	29.0
A2	6	22.7	4.3	19.0	31.0
F	5	23.6	8.6	19.0	39.0
G	6	20.8	3.2	17.0	26.0
H	5	21.4	3.0	18.0	26.0
I	4	23.5	6.0	18.0	32.0
J	2	24.0	5.7	20.0	28.0
K	2	25.5	7.8	20.0	31.0
Brooks	6	27.2	5.4	23.0	37.0
OU	30	22.6	5.0	17.0	39.0
15 ED	8	24.5	6.2	19.0	37.0
5 on/2 off	2	25.0	2.8	23.0	27.0
2 on/5 off	2	26.5	3.5	24.0	29.0
7 x 2 per day	15	22.8	5.9	17.0	39.0
15 x 2 (EOD)	9	22.3	4.4	18.0	32.0
3-day break	20	22.8	4.4	17.0	32.0
5-day break	16	24.1	6.3	18.0	39.0

Males

	Count	Mean Age	Std. Dev. Age	Min Age	Max Age
	65	27.5	9.0	18.0	52.0
A1	4	42.3	11.6	25.0	50.0
B	2	32.5	9.2	26.0	39.0
C	2	49.5	3.5	47.0	52.0
D	4	35.5	8.6	26.0	45.0
E	3	33.3	10.1	24.0	44.0
A2	14	24.9	6.7	19.0	44.0
F	5	24.2	3.1	21.0	29.0
G	4	20.8	3.2	18.0	24.0
H	4	25.0	4.2	21.0	30.0
I	6	24.3	4.9	18.0	32.0
J	8	25.3	3.8	20.0	30.0
K	9	23.8	7.7	19.0	44.0
Brooks	15	38.3	10.0	24.0	52.0
OU	50	24.3	5.5	18.0	44.0
15 ED	20	29.1	10.4	19.0	50.0
5 on/2 off	6	40.2	9.9	26.0	52.0
2 on/5 off	3	33.3	10.1	24.0	44.0
7 x 2 per day	26	23.8	5.3	18.0	44.0
15 x 2 (EOD)	10	24.6	4.4	18.0	32.0
3-day break	39	27.2	10.2	18.0	52.0
5-day break	26	28.1	7.0	20.0	45.0

APPENDIX B

TASK DESCRIPTIONS

BRIEF DESCRIPTION OF PAWS

Mood Scale II. The Mood Scale II is a variation of the Profile of Mood States (POMS; McNair, Lorr, and Droppleman, 1971). The Mood Scale II has 36 adjectives which address Activity, Happiness, Depression, Anger, Fatigue, and Fear.

Tracking. One of the primary potential effects of microgravity is a disruption of visual-motor coordination due to disturbances in the sensory input and motor output channels. One of the candidate tracking algorithms is the Crossover Model developed by McRuer and Jex (1967) and validated by DOD, NASA, and FAA in numerous studies. The task requires that the subject maintain an unstable target in the center of a horizontal line by manipulating a control device to nullify the input disturbance. An instability parameter (lambda) is used to control the difficulty of the task. In this study, different tracking algorithms and rules for specifying the lambda parameter will be evaluated in order to specify the tracking test's final configuration.

Matrix Rotation Task. This test, developed by Damos and Lyall (1984), uses 100 basic patterns. Each pattern is a 5 by 5 matrix with five illuminated cells that have been selected at random. At the beginning of the trial, the subject sees a pattern. After studying the pattern, the subject presses a response key. The pattern is immediately erased and a new one is presented. The subject must decide as quickly as possible if the new pattern is identical except for a rotation to the preceding pattern. The subject then presses one key for "same" or another key for "different." As soon as the response is made, a third pattern appears. The subject must now compare the new pattern to the immediately preceding pattern, etc. For "same" responses, the two patterns are not presented in the same orientation; the second pattern is always rotated either 90 degrees to the left or 90 degrees to the right relative to the preceding pattern. Both mean reaction time for correct responses and percentage correct are used as dependent measures.

Sternberg Memory Search. The general Sternberg paradigm (Sternberg, 1969) requires that subjects respond as rapidly and accurately as possible to visually presented letters. At the beginning of the test, a set of letters drawn randomly from a restricted alphabet is presented to the subject for memorization. The set of letters (positive set) stays on the screen for a maximum of five seconds, then the screen is cleared and a series of single test letters is presented. If the presented letter matches one of the letters in the previously memorized positive set, the subject responds "same" (key press). If a different letter appears (negative set), then the subject responds "different" (key press), indicating a non-matching letter was presented. The Sternberg task included in this version of the UTC-PAB uses a set size of four letters that are changed after each block of trials. Thus, a letter can be a target in one session and a distracter in another.

Continuous Recognition. One critical aspect of higher cognitive function is the ability to maintain attention and to carry out repetitive cognitive processes over some period of time. In many ways, such activities encompass those which were traditionally referred to as "vigilance." However, they add the dimension of active processing of information, rather than simple monitoring. One task that appears to capture the performance elements above is Continuous Recognition (Hunter, 1975; Shingledecker, 1984). In this task, the subject sees two numbers, one above the other. The subject is to remember the bottom number. When the next two numbers appear, the task is to determine if the new top number is the same as the previous bottom number. However, before responding, one must note the new bottom number because as soon as a response is made, the numbers are replaced by a new pair. Thus, the subject must not only exercise very short-term memory, but more importantly, must inhibit the response until the new bottom number is committed to memory. The appropriate strategy is to develop a set pattern of observing, memorizing, observing, comparing, and responding. This sequence is different enough from that required by most routine tasks in that it requires constant attention allocation. Even brief lapses result in errors. The task can be made more difficult by requiring that the subject remember and respond to numbers further removed from the immediately preceding one (e.g., two-back or even three-back), thus imposing a much higher load on immediate memory.

Performance Switching Task - Manikin and Mathematical Processing. Time-sharing, as explained above in the Dual Task, is different from attention switching, another required attentional process that could be affected by microgravity. Astronauts must make rapid shifts in attentional focus, as well as in the skills required to respond to a change in task demands. This externally-directed behavior defies automaticity in any true sense, since it must be flexible enough to respond to unusual demands. Thus, a test is needed to probe the subject's ability to shift attention and resource allocation in response to rapidly changing and unpredictable external demands. Such a procedure has been created that uses two tasks currently in the UTC-PAB.

In this procedure, the subject has two distinct and discrete tasks to perform. One is a spatially-based task, and the other is a mathematically-based test. Each of these appears, side-by-side, simultaneously on the screen on every trial. However, an indicator appears at the same time directing the subject to the task that is "active" (i.e., must be responded to). The subject must make an exclusive response to the active task, where reaction time and percent-correct data are obtained only for that task. The switching from task to task for each trial is random (within constraints). Therefore, the subject must remember to watch the indicator on each trial, allocate the appropriate resources to respond to that trial, and then make the appropriate response. This paradigm provides a test of the switching skills described above. Findings and results using this test can be found in O'Donnell (1991).

The two tests selected to exercise this paradigm are the Manikin test and the Mathematical Processing test. The Manikin test has a long history of use (Benson & Gedye, 1963; Reader, Benel, & Rahe, 1981; Schlegel and Storm, 1983) and is presented in a wide variety of formats by military psychologists (Miller, Takamoto, Bartel, & Brown, 1985). As implemented in this microgravity experiment, a manikin "stick figure" is presented facing either forward or backward. In addition, the figure can be either upright or upside-down. The figure is also standing on a box and inside the box is either a rectangle or a circle. In the figure's two hands are a rectangle and a circle. The subject's task is to note which symbol is inside the box, and then to determine which of the manikin's hands is holding the designated symbol. The subject then presses the left or right of two keys corresponding to the manikin's left or right hand.

The Mathematical Processing test is based on similar tasks described by Perez et al. (1987). It presents two single-digit numbers that must be added or subtracted. If the answer is greater than 5, one response is given. If the answer is less than 5, another response is required. This task has been reported by Shingledecker (1984) to be a relatively pure index of mathematical functioning.

Dual Task - Tracking and Sternberg Memory Search. One of the most critical and potentially sensitive higher cognitive functions that might be affected by microgravity is the ability of the subject to allocate attentional resources among several tasks. To investigate this, the present study will use the time-sharing paradigm that has been well studied in cognitive psychology (Damos & Wickens, 1980; O'Donnell & Eggemeier, 1986; Damos, 1991). The specific form of this paradigm will be the dual task included in the UTC-PAB. This consists of the Sternberg task and the Tracking task being presented simultaneously. In this implementation of the Dual Task, the tracking task is presented in the middle of the screen and the letters of the Sternberg task appear in a fixed location directly above the center null point. The target of the compensatory tracking task moves laterally. One memory set will be used, consisting of four letters. Due to the nature of the dual task, the same "fixed set" procedure as in the single Memory Search task will be used. That is, only one positive memory set is presented with several probe letters for each daily session. For a recent study discussing the implementation of the dual task when investigating the effects of antihistamines on military weapon system controllers, see Nesthus, Schiflett, Eddy, and Whitmore (1991).

Fatigue Scale. The fatigue scale is a scale designed to assess the level of fatigue experienced by the subjects. The subject simply responds to the scale by selecting the statement that best describes the level of fatigue at that moment. The scale is presented on the screen with the following statements:

Choose one of the seven statements below that best describes your present feeling.

HOW DO YOU FEEL RIGHT NOW?

1. Fully Alert; Wide Awake; Extremely Peppy.
2. Very Lively; Responsive, But Not at Peak.
3. Okay; Somewhat Fresh.
4. A Little Tired; Less Than Fresh.
5. Moderately Tired; Let Down.
6. Extremely Tired; Very Difficult to Concentrate.
7. Completely Exhausted; Unable to Function Effectively; Ready to Drop.

References

Benson, A.J. and Gedye, J.L. (1963). Logical processes in the resolution of orienting conflict. RAF Rpt. 259, Farnborough, UK, Royal Air Force Institute of Aviation Medicine.

Damos, D.L. and Lyall, E.A. (1984). The effect of asymmetric transfer on dual-task assessment of voice technology. Proceedings of the 28th Annual Meeting of the Human Factors Society, Santa Monica, CA: Human Factors Society.

Damos, D.L. and Wickens, C.D. (1980). The identification and transfer of timesharing skills. Acta Psychologica, 46, 15-39.

Damos, D.L. (1991). Multiple Task Performance. London, UK, Taylor & Francis, LTD.

Hunter, D.R. (1975). Development of an enlisted psychomotor and perceptual test battery. AFHRL-TR-75-60, Brooks AFB, TX: Air Force Human Resources Laboratory.

McNair, D.M., Lorr, M., and Droppleman, L.F. (1971). *Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.

McRuer, D.T. and Jex. H.R. (1967). A review of quasi-linear pilot models. IEEE Transactions on Human Factors in Electronics, 8, 231-249.

Miller, J.C., Takamoto, G.M., Bartel, G.M., and Brown, M.D. (1985). Psychophysiological correlates of long-term attention to complex tasks. Behavior Research Methods, Instruments, and Computers, 17(2), 186-190.

Nesthus, T.E., Schiflett, S.G., Eddy, D.R., and Whitmore, J.N. (1991). Comparative effects of antihistamines on aircrew performance of simple and complex tasks under sustained operations. AL-TR-91-104, Brooks AFB, TX: Armstrong Laboratory, Crew Technology Division.

O'Donnell, R.D. (1991). Scientific validation of the Novascan (tm) tests: Theoretical basis and initial validation studies. NTI Report to Nova Technology, Inc., 19460 Shenango Drive, Tarzana, CA: NTI, Incorporated.

O'Donnell, R.D. and Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff et al. (eds), Handbook of Perception and Human Performance, Vol II, New York: Wiley, 42-1 to 42-49.

Perez, W.A., Masline, P.J., Ramsey, F.R., and Urban, K.E., (1987). Unified Tri-Services Cognitive Performance Assessment Battery: Review and Methodology. AAMRL-TR-87, Wright-Patterson AFB, OH: Armstrong Aerospace Medical Research Laboratory.

Schlegel, R.E. and Storm, W.F. (1983). Speed-accuracy tradeoffs in spatial orientation information processing. Proceedings of the 27th Annual Meeting of the Human Factors Society, Santa Monica, CA: Human Factors Society.

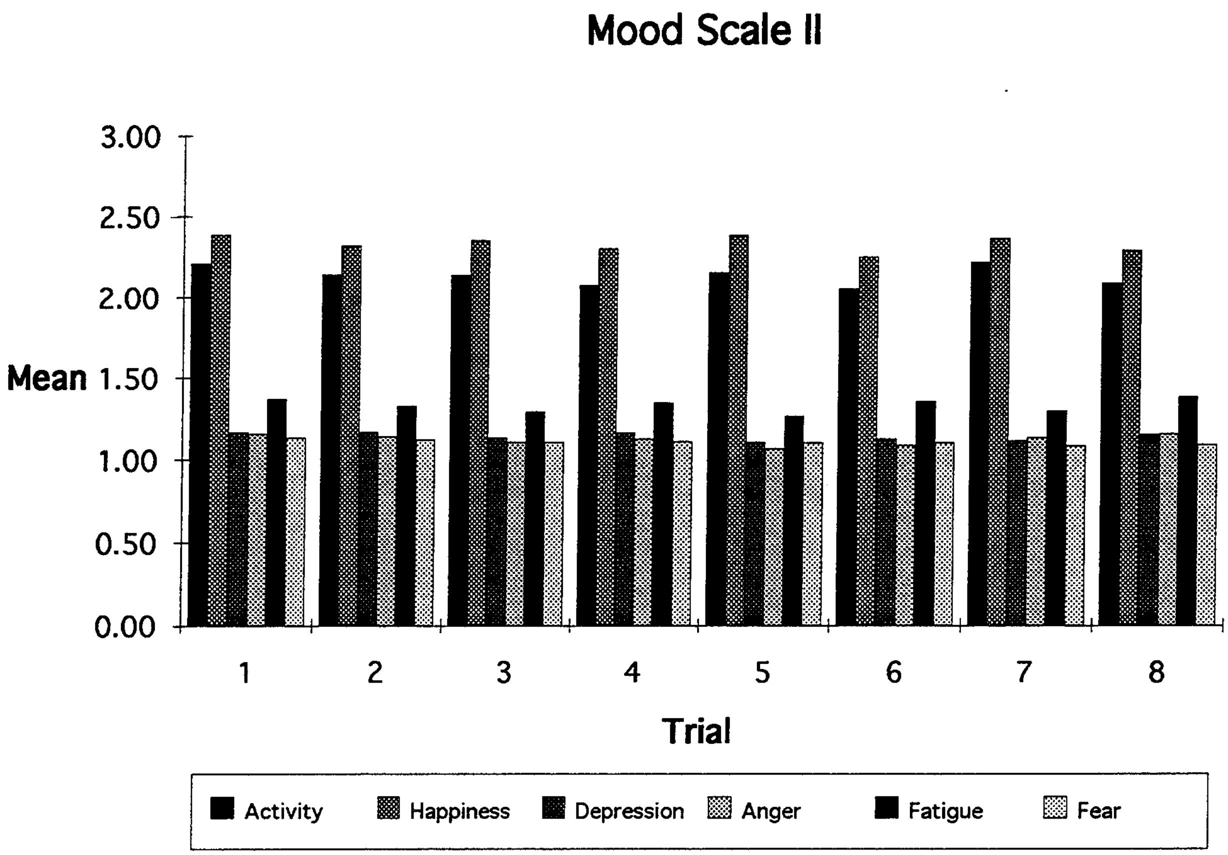
Reader, D.C., Benel, R.A., and Rahe, A.J. (1981). Evaluation of a manikin psychomotor task. USAFSAM-TR-81-10, Brooks AFB, TX: USAF School of Aerospace Medicine.

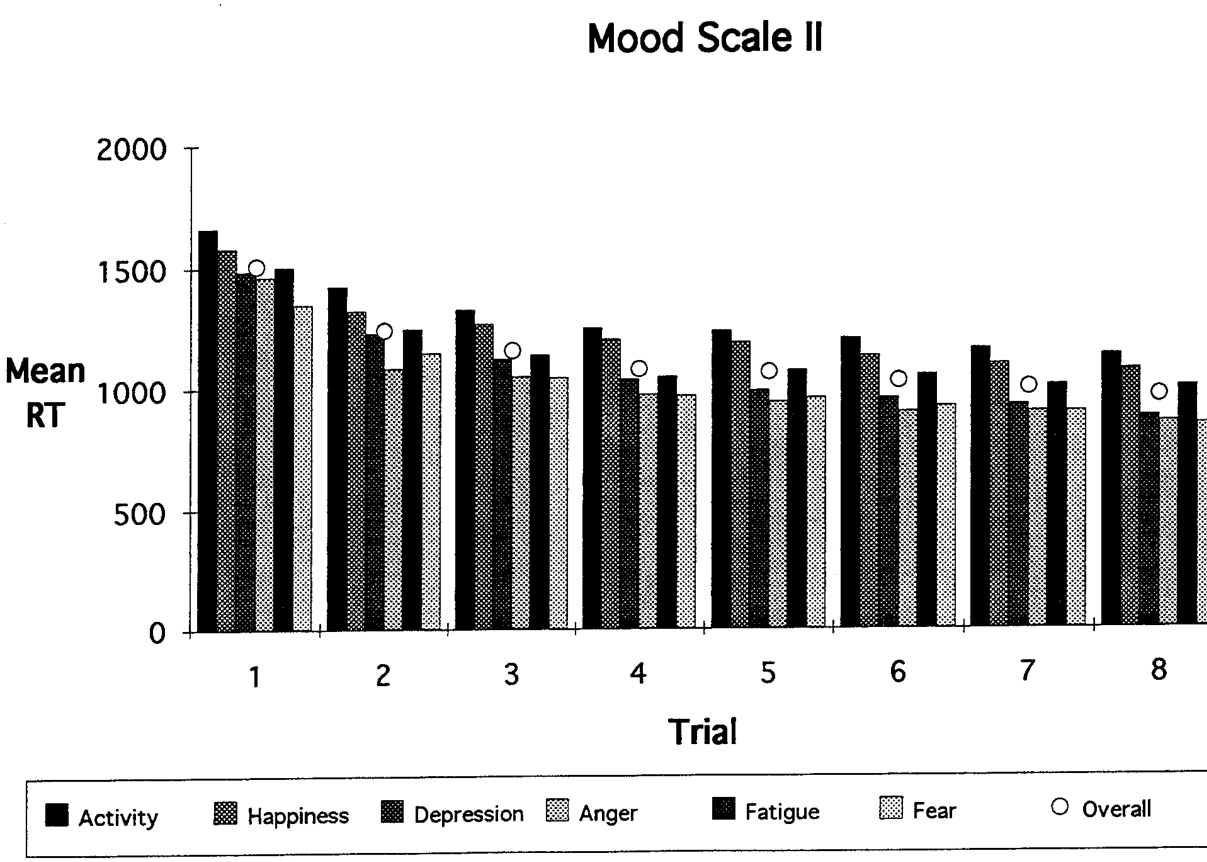
Shingledecker, C.A. (1984). A task battery for applied human performance assessment research. AFAMRL-TR-84-071, Wright-Patterson AFB, OH: Air Force Aerospace Medical Research Laboratories.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. Acta Psychologica, 30, 276-315.

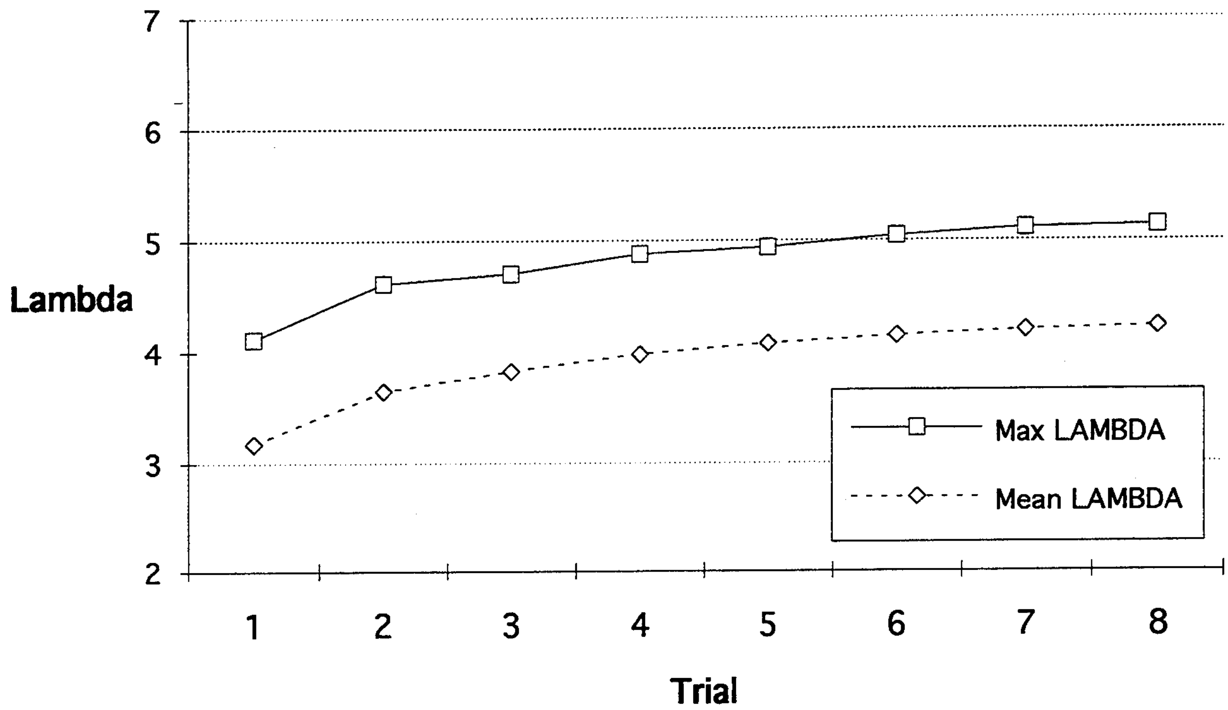
APPENDIX C

TRAINING DATA

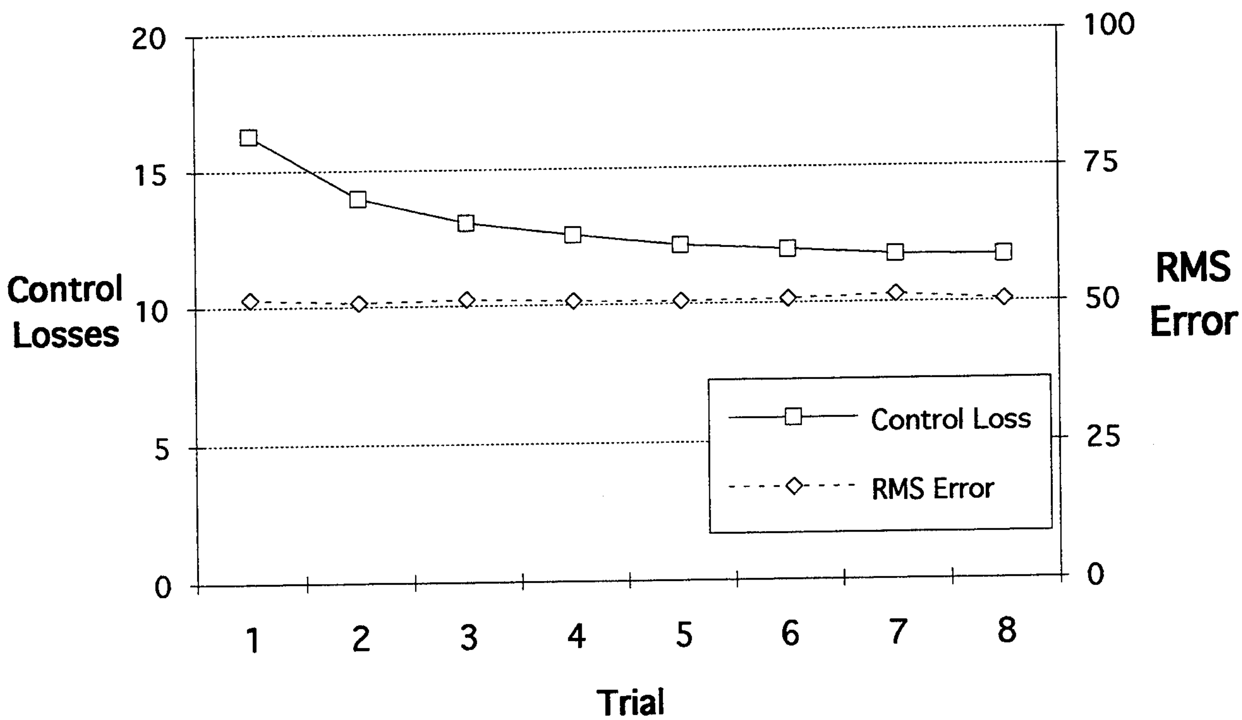

Mood Scale II
3.00
2.50
2.00
Mean 1.50
1.00
0.50
0.00
1
2
3
4
5
6
7
8
Trial
Activity
Happiness
Depression
Anger
Fatigue
Fear

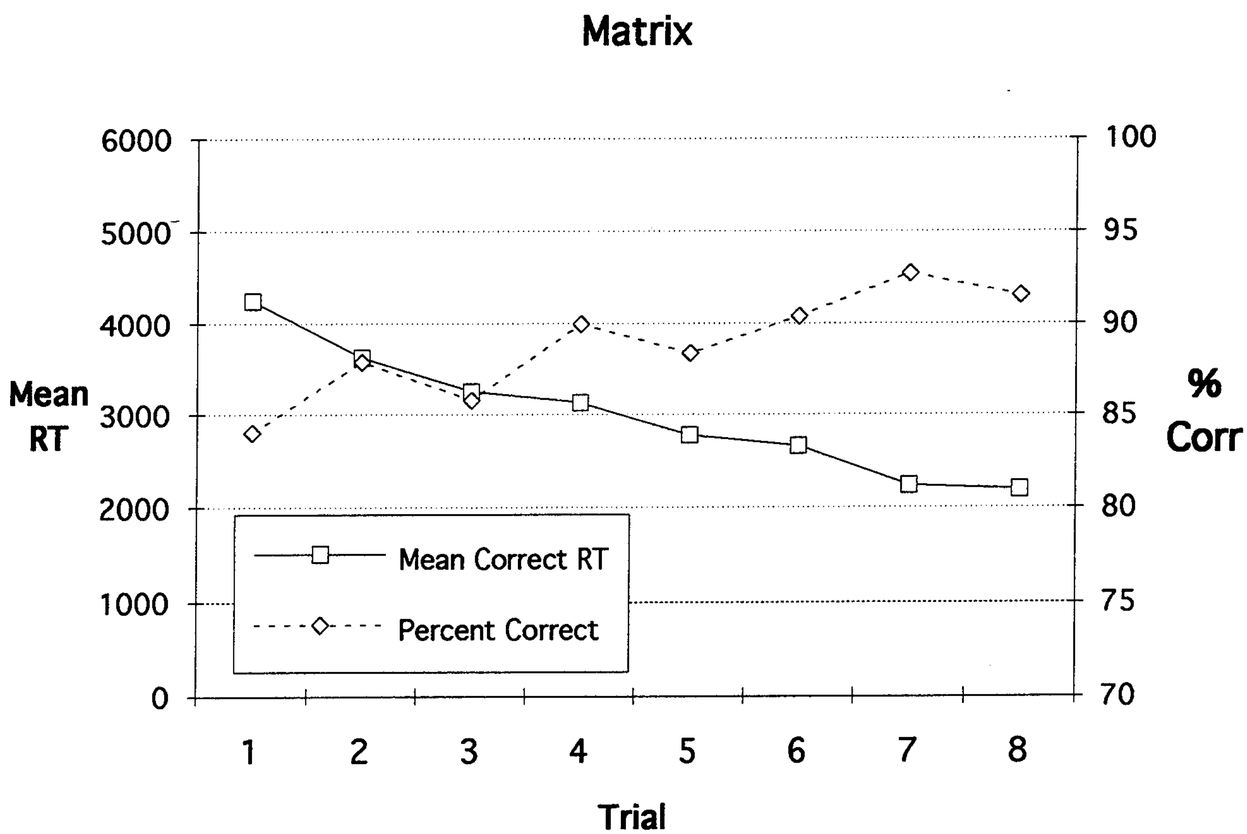

Mood Scale II
2000
1500
Mean RT
1000
500
0
1
2
3
4
5
6
7
8
Trial
Activity
Happiness
Depression
Anger
Fatigue
Fear
Overall

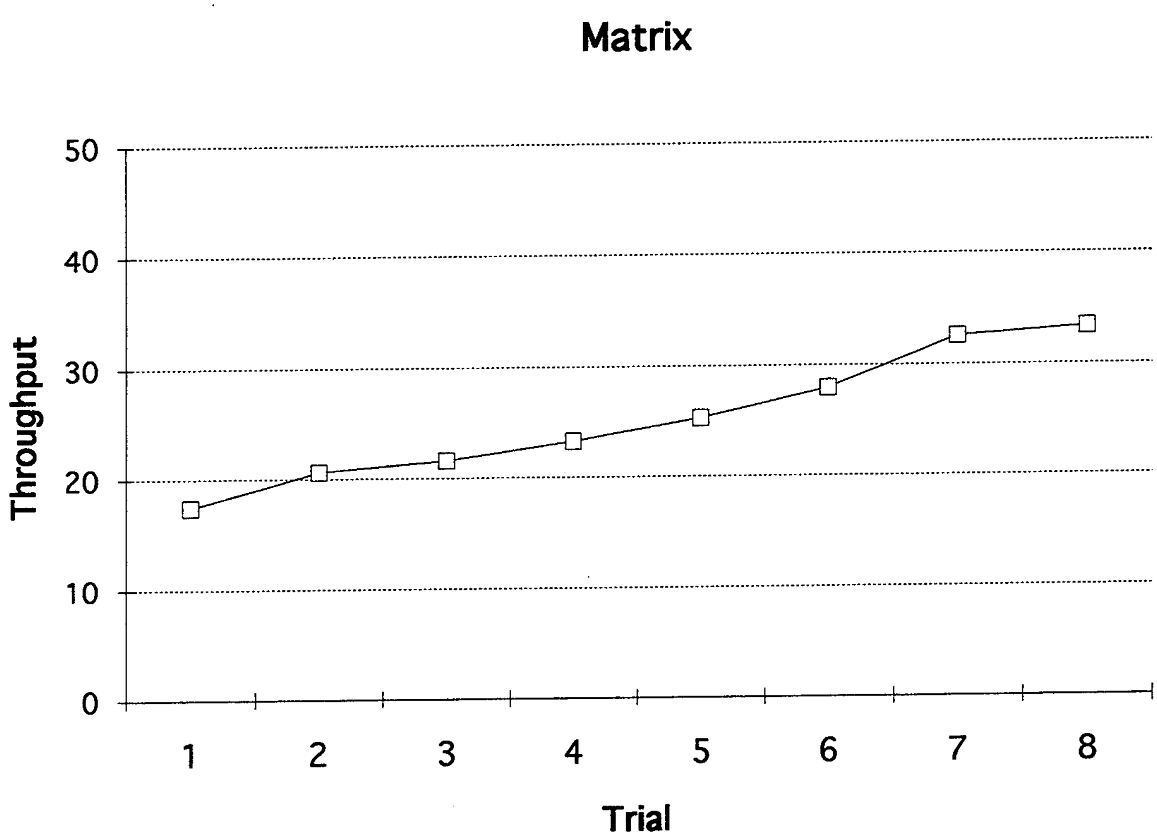
Critical Tracking



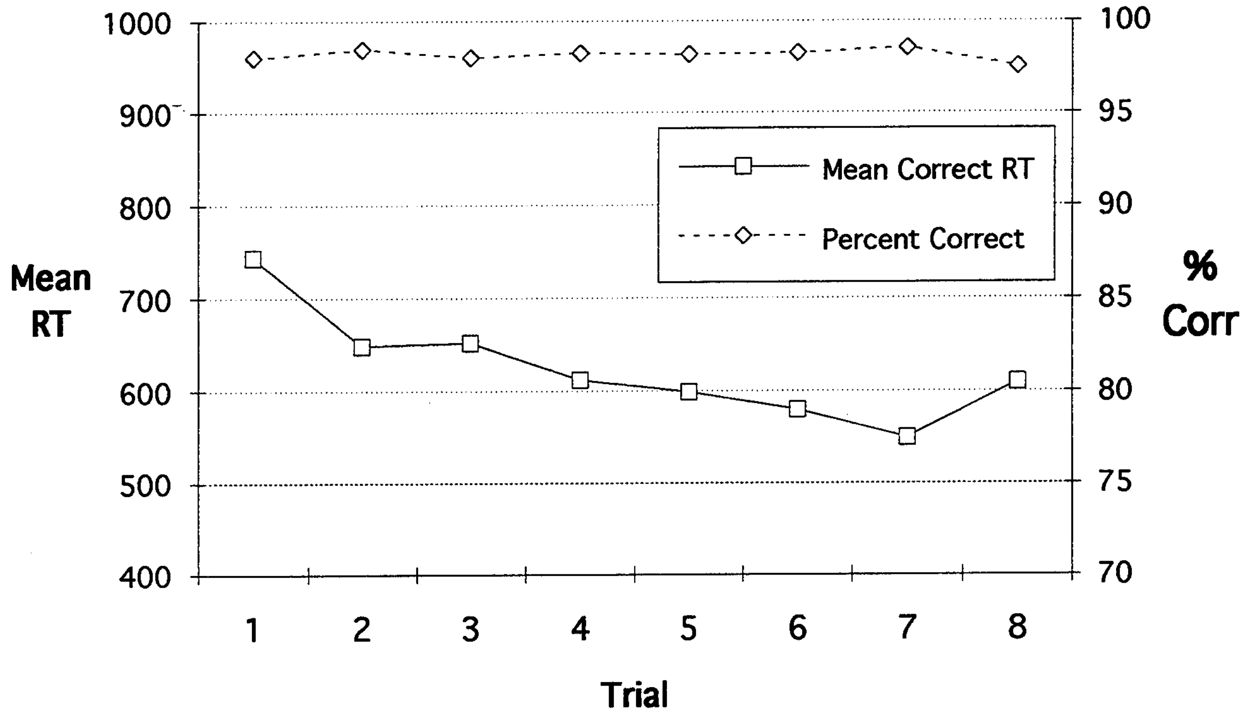
Critical Tracking



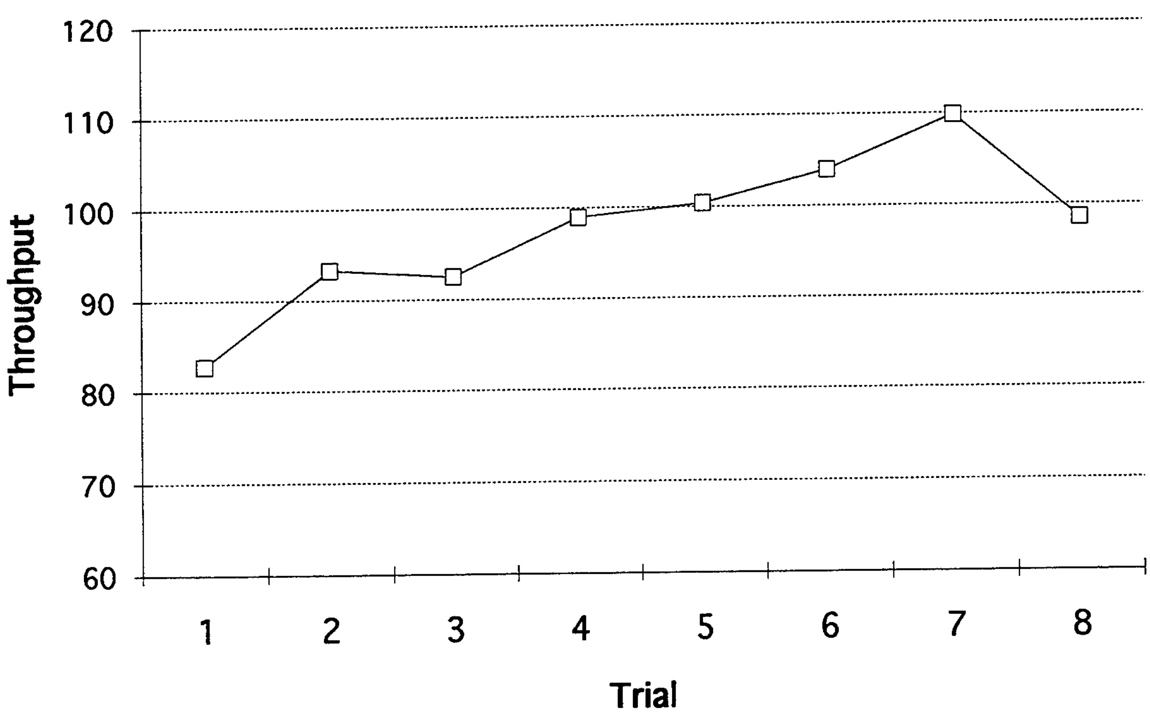

Matrix
6000
5000
4000
3000
2000
1000
0
Mean
RT
100
95
90
85
80
75
70
%
Corr
Mean Correct RT
Percent Correct
1
2
3
4
5
6
7
8
Trial


Matrix
50
40
30
20
10
0
Throughput
1
2
3
4
5
6
7
8
Trial

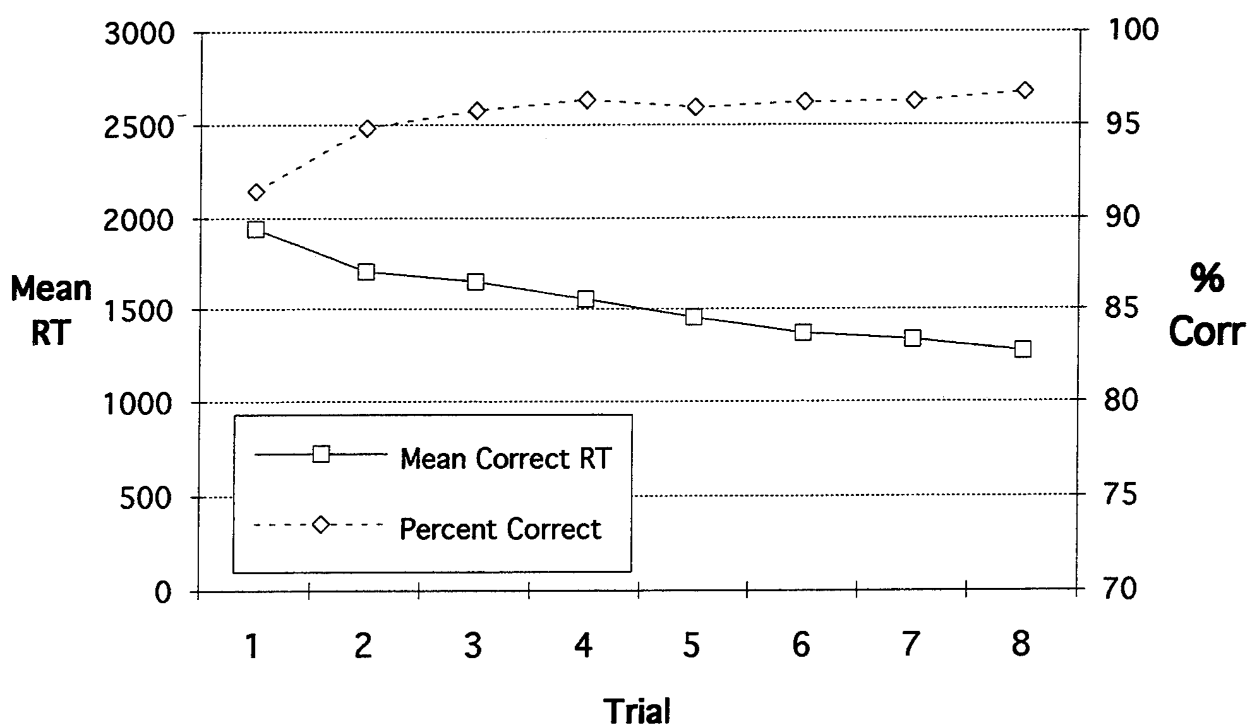
Memory Search



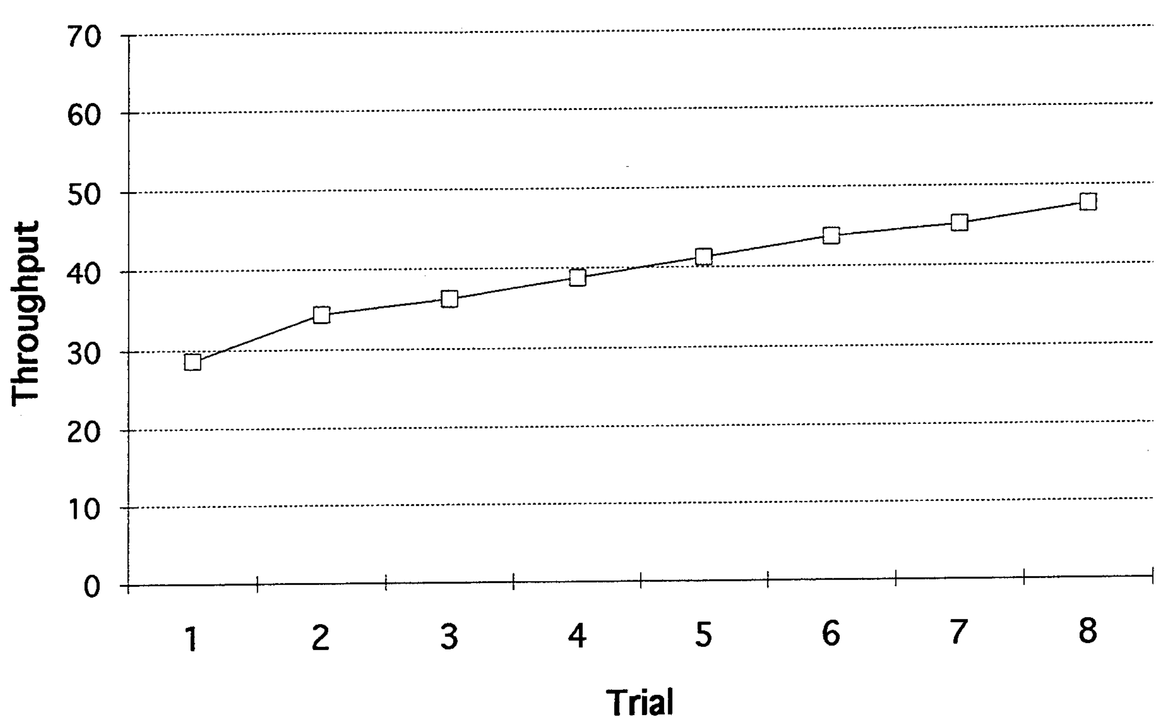
Memory Search

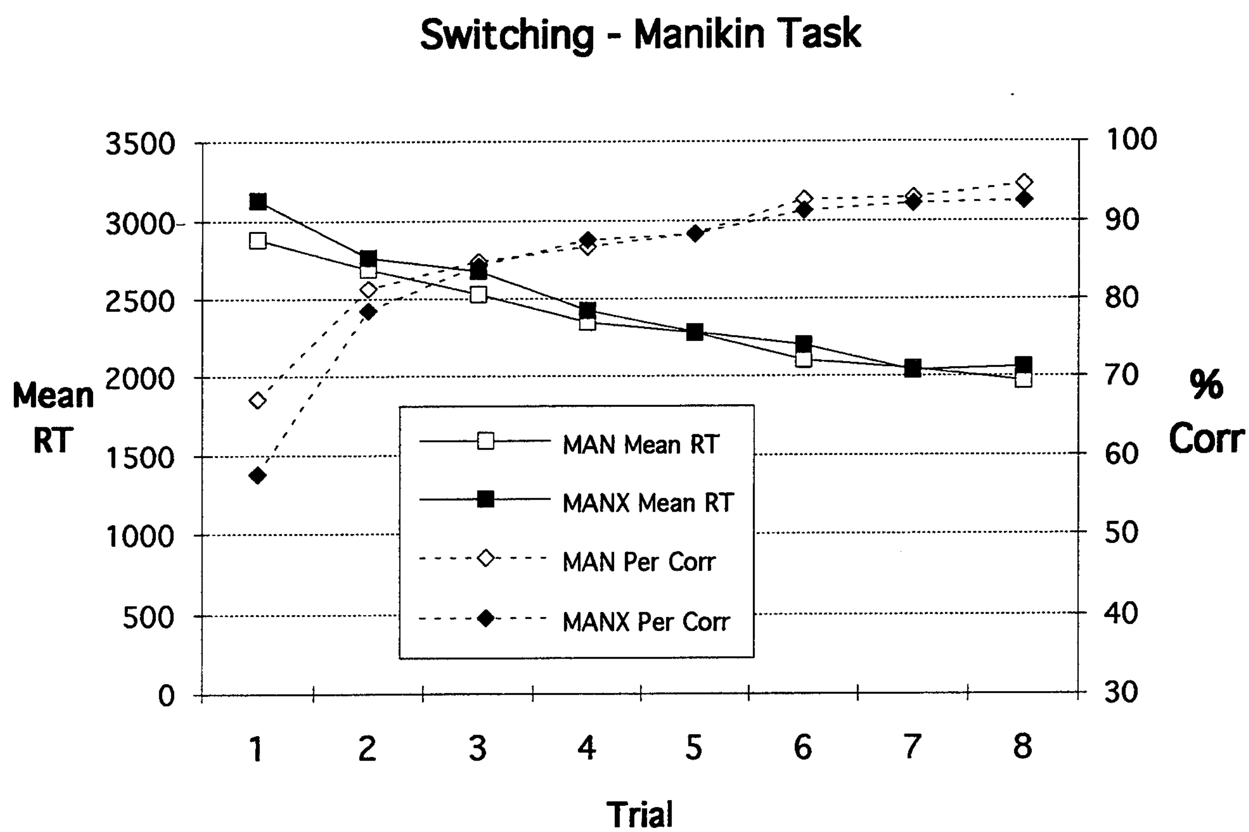


Continuous Recognition

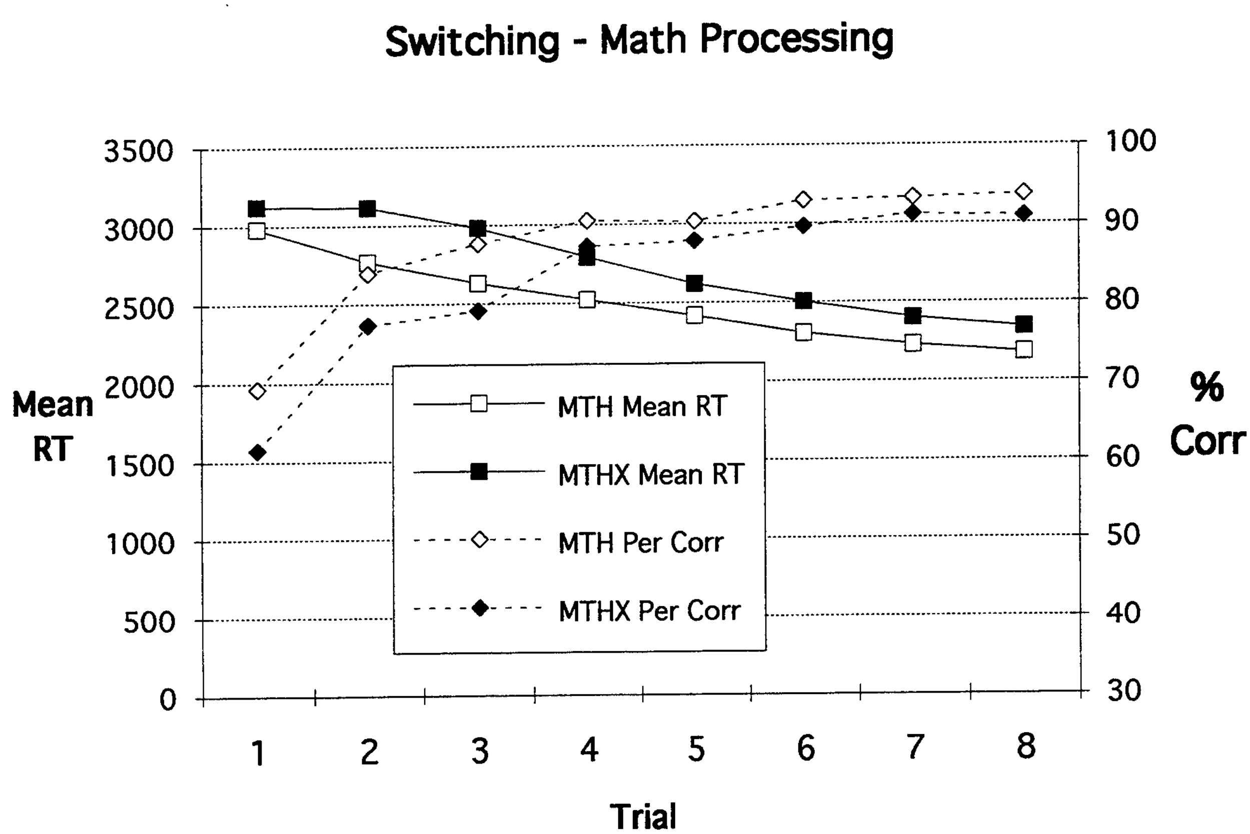


Continuous Recognition

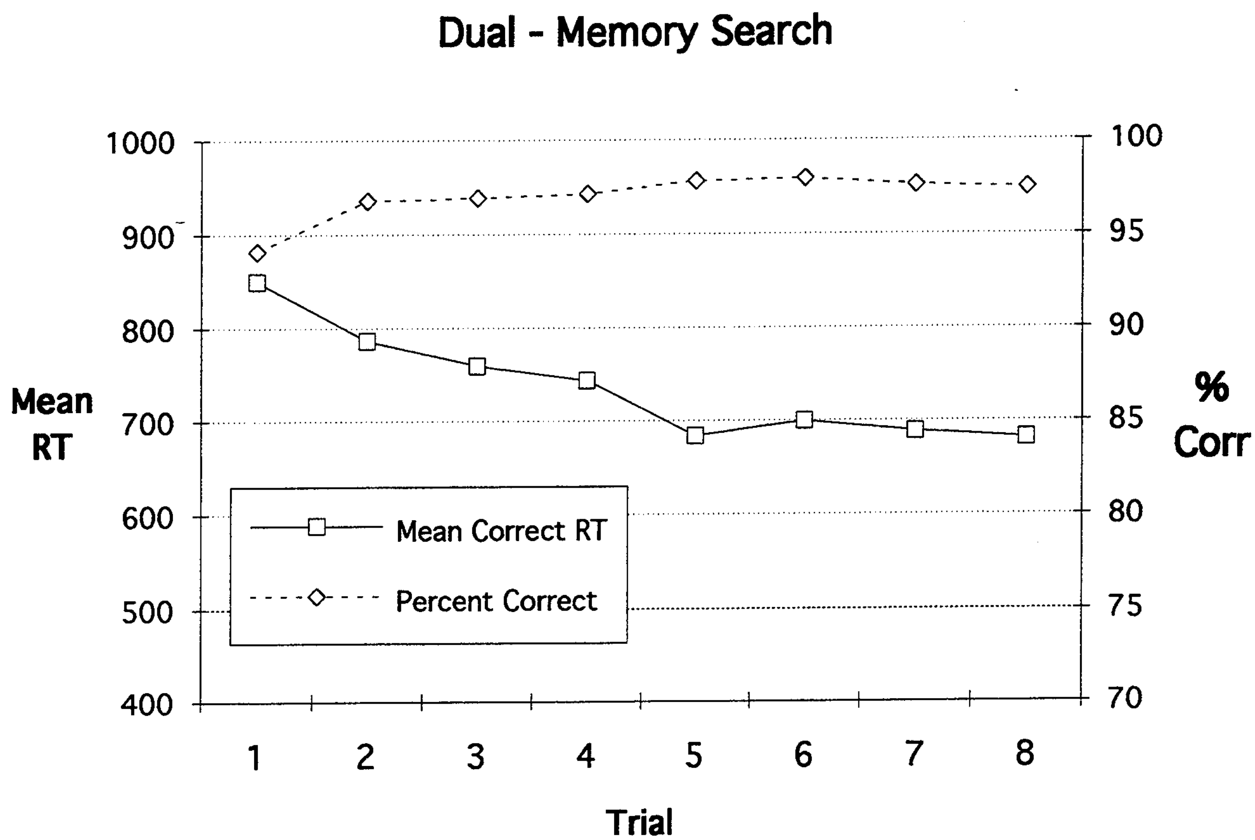


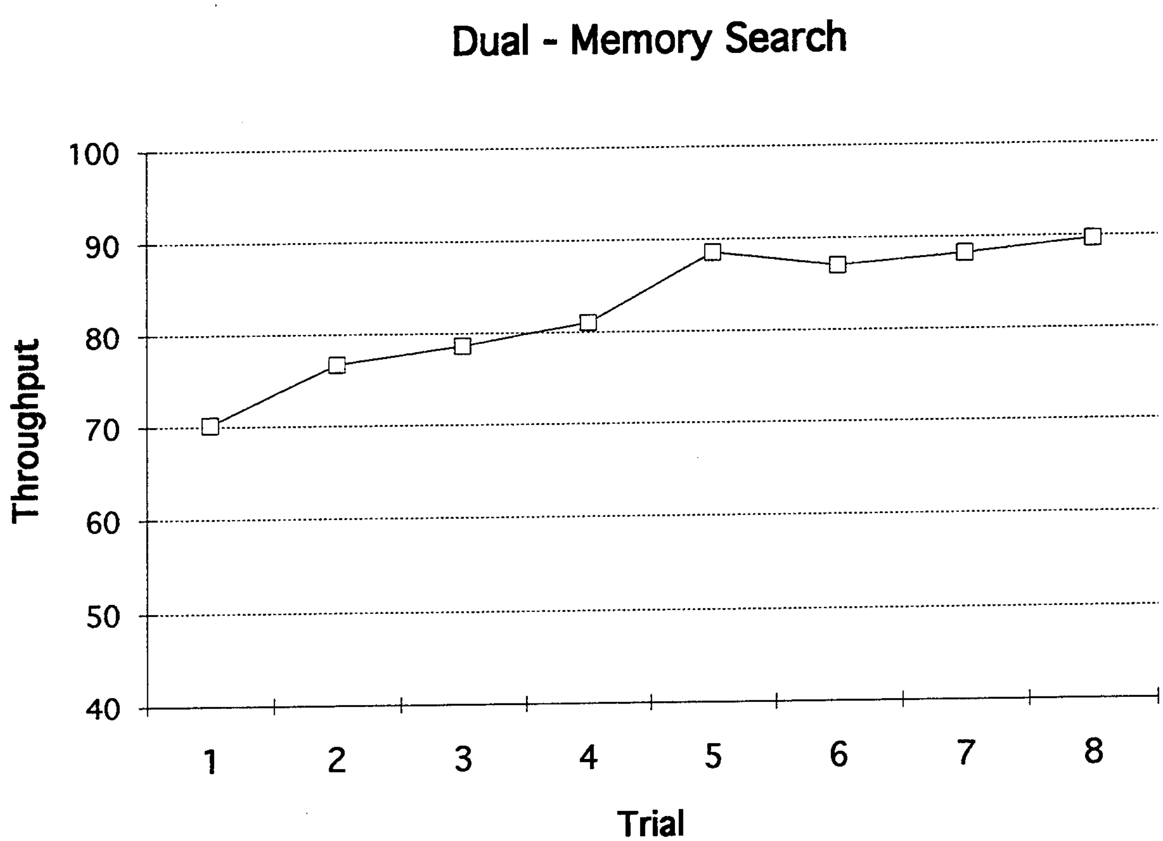


Switching - Manikin Task
Mean RT
% Corr
3500
3000
2500
2000
1500
1000
500
0
100
90
80
70
60
50
40
30
1
2
3
4
5
6
7
8
Trial
MAN Mean RT
MANX Mean RT
MAN Per Corr
MANX Per Corr

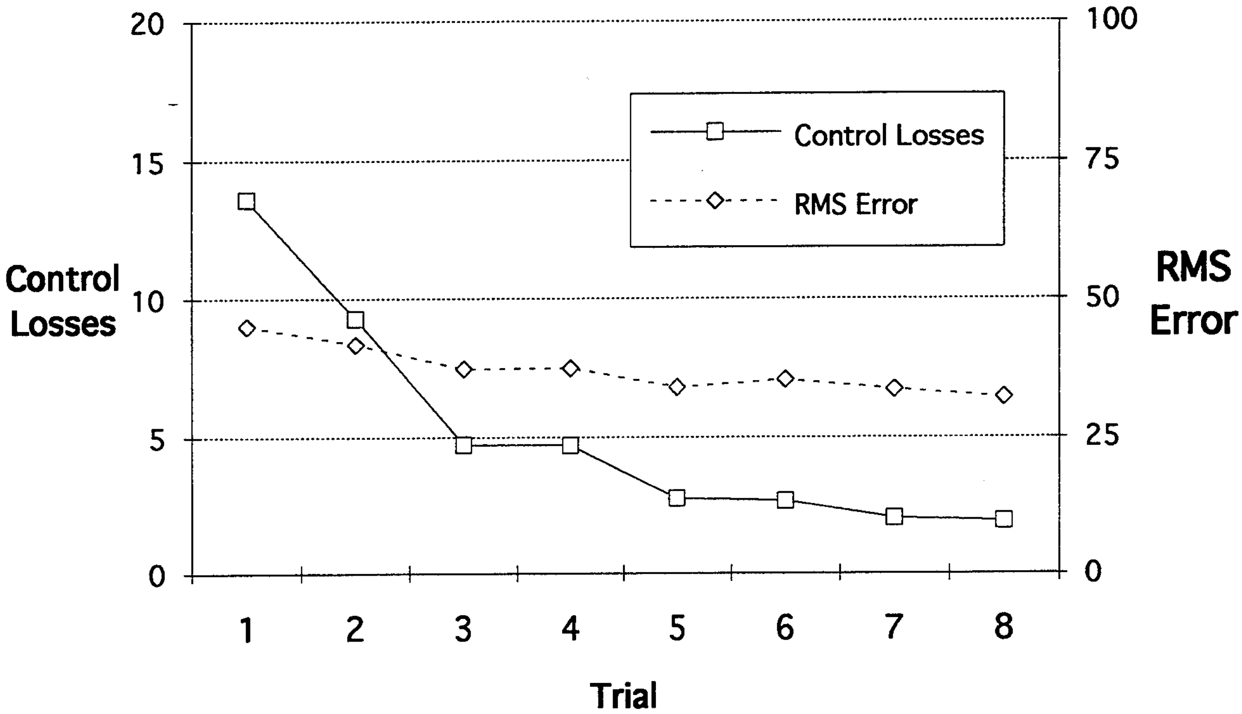


Switching - Math Processing
Mean RT
% Corr
3500
3000
2500
2000
1500
1000
500
0
100
90
80
70
60
50
40
30
1
2
3
4
5
6
7
8
Trial
MTH Mean RT
MTHX Mean RT
MTH Per Corr
MTHX Per Corr

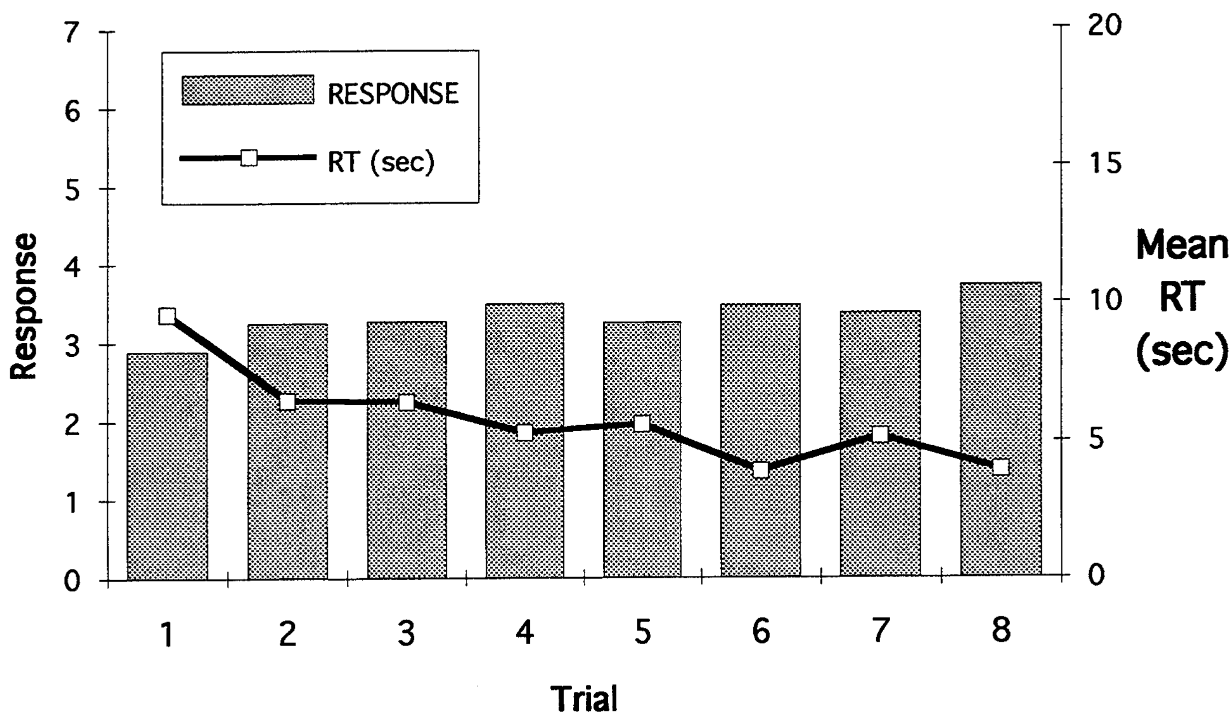

Dual - Memory Search
1000
900
800
700
600
500
400
Mean
RT
100
95
90
85
80
75
70
%
Corr
Mean Correct RT
Percent Correct
1
2
3
4
5
6
7
8
Trial


Dual - Memory Search
100
90
80
70
60
50
40
Throughput
1
2
3
4
5
6
7
8
Trial

Dual - Tracking



Fatigue Scale



APPENDIX D

TEST SITE DATA
(BROOKS AFB vs. UNIVERSITY OF OKLAHOMA)

	Group	Train	Start	End	MD-1	MD-2	MD-3	MD-4	MD-5
TRK LM	Brooks	5.1	5.2	6.1	5.9	5.9	6.0	6.0	6.0
	OU	5.1	5.1	5.7	5.7	5.7	5.7	5.7	5.8
TRK ML	Brooks	-	-	-	-	-	-	-	-
	OU	4.2	4.2	4.8	4.8	4.9	4.9	4.9	5.0
TRK CL	Brooks	11.6	11.4	9.4	9.4	9.3	9.4	9.0	9.1
	OU	11.8	11.7	10.1	10.0	10.1	10.0	10.1	9.7
TRK RMS	Brooks	48.6	46.9	47.0	48.5	47.3	48.2	48.0	48.0
	OU	51.7	51.5	51.4	50.5	50.8	50.8	50.7	50.9
MTX RT	Brooks	2713	2264	1531	1529	1408	1358	1277	1177
	OU	2026	1787	1335	1348	1332	1304	1288	1212
MTX PC	Brooks	93.7	88.4	93.9	91.6	91.7	92.8	91.7	90.6
	OU	91.9	89.3	92.8	91.4	92.0	92.9	91.5	92.2
MTX TP	Brooks	29.9	30.7	45.9	45.9	48.5	50.1	50.6	52.7
	OU	34.0	34.9	48.5	47.2	48.1	49.5	50.1	51.8
STN RT	Brooks	587	594	499	533	509	503	542	521
	OU	579	592	533	551	540	534	568	554
STN PC	Brooks	96.8	97.2	97.8	97.0	96.9	97.9	95.8	96.4
	OU	98.2	98.9	98.1	98.2	98.4	98.4	97.8	98.0
STN TP	Brooks	100.4	101.2	119.1	110.8	116.0	118.5	107.9	112.5
	OU	104.6	102.5	113.5	109.7	112.7	113.1	106.6	109.1
CRC RT	Brooks	1462	1355	998	955	908	914	922	876
	OU	1264	1256	965	950	938	931	918	906
CRC PC	Brooks	92.5	93.3	93.3	93.5	93.1	92.4	92.9	93.0
	OU	97.2	97.4	97.1	97.1	97.2	97.8	97.2	96.5
CRC TP	Brooks	39.0	42.7	59.0	61.8	64.6	63.2	63.9	66.6
	OU	48.0	48.6	63.0	63.6	64.5	65.6	65.7	66.7

	Group	Train	Start	End	MD-1	MD-2	MD-3	MD-4	MD-5
MAN RT	Brooks	2167	2184	1530	1538	1475	1493	1424	1460
	OU	1970	1981	1474	1526	1489	1469	1453	1414
MAN PC	Brooks	92.9	91.9	96.9	96.5	97.1	96.1	96.9	97.4
	OU	94.3	93.9	98.1	97.9	98.3	98.2	97.9	98.3
MAN RTX	Brooks	-	-	-	-	-	-	-	-
	OU	2047	2073	1585	1661	1653	1612	1615	1576
MAN PCX	Brooks	-	-	-	-	-	-	-	-
	OU	92.4	93.1	97.7	97.5	98.2	97.5	97.2	97.9
MTH RT	Brooks	2219	2281	1867	1859	1819	1857	1800	1718
	OU	2201	2249	1839	1913	1897	1786	1824	1745
MTH PC	Brooks	93.9	92.3	95.6	94.6	92.8	94.1	93.1	95.5
	OU	93.3	90.5	96.7	96.1	97.0	96.8	96.6	96.7
MTH RTX	Brooks	-	-	-	-	-	-	-	-
	OU	2369	2342	2041	2143	2079	1965	2032	1946
MTH PCX	Brooks	-	-	-	-	-	-	-	-
	OU	91.0	90.6	96.5	96.1	96.5	97.3	96.8	96.8
DUL CL	Brooks	3.1	3.0	7.5	9.5	6.8	8.5	5.9	5.9
	OU	1.7	1.8	11.9	11.3	10.0	9.9	9.7	10.0
DUL RMS	Brooks	30.2	23.9	45.4	48.9	46.7	46.5	47.4	44.5
	OU	33.2	32.4	52.6	53.0	49.6	51.7	50.2	49.5
DUL RT	Brooks	687	674	688	683	680	707	658	688
	OU	686	692	702	675	661	688	671	697
DUL PC	Brooks	97.0	96.4	96.5	96.4	97.0	96.7	96.8	94.7
	OU	97.6	97.9	97.4	98.2	97.7	97.9	98.0	96.9
DUL TP	Brooks	86.7	88.0	88.7	90.2	92.5	90.7	94.4	88.9
	OU	89.2	87.8	87.8	91.0	93.1	90.1	91.9	87.1
FAT RP	Brooks	3.8	2.9	2.5	3.2	2.7	3.3	2.8	2.9
	OU	3.5	3.3	3.3	3.4	3.5	3.2	3.4	2.9
FAT RT	Brooks	-	-	-	-	-	-	-	-
	OU	4.3	5.0	2.7	3.2	2.2	3.2	2.3	2.4

APPENDIX E

TEST GROUP DATA

	Group	Train	Start	End	MD-1	MD-2	MD-3	MD-4	MD-5
TRK LM	A1	5.3	5.2	6.2	6.1	6.0	6.0	6.1	6.3
	B	5.0	5.2	5.9	5.7	5.6	6.0	5.9	5.7
	C	4.8	4.7	5.6	5.5	5.7	5.7	5.7	5.7
	D	5.4	5.6	6.6	6.4	6.5	6.4	6.3	6.3
	E	5.3	5.5	5.4	5.5	5.5	6.1	5.7	5.8
	A2	5.3	5.2	5.8	5.9	5.9	5.8	5.9	6.0
	F	5.2	5.1	5.7	5.8	5.7	5.8	5.6	6.0
	G	5.0	4.9	5.6	5.5	5.8	5.8	5.7	5.9
	H	4.7	4.8	5.4	5.5	5.2	5.5	5.5	5.4
	I	4.8	4.9	5.4	5.4	5.4	5.4	5.3	5.4
	J	5.4	5.3	5.9	6.0	5.9	6.0	6.0	5.9
	K	5.2	5.3	5.7	5.7	5.9	5.7	5.9	6.0
TRK ML	A1	-	-	-	-	-	-	-	-
	B	-	-	-	-	-	-	-	-
	C	-	-	-	-	-	-	-	-
	D	-	-	-	-	-	-	-	-
	E	-	-	-	-	-	-	-	-
	A2	4.3	4.4	4.9	5.0	5.0	4.9	5.1	5.2
	F	4.1	4.1	4.7	4.9	5.0	4.9	4.7	4.9
	G	4.1	4.1	4.7	4.7	4.9	5.0	4.8	5.0
	H	3.8	3.9	4.5	4.5	4.5	4.7	4.7	4.5
	I	3.9	4.0	4.6	4.5	4.5	4.6	4.5	4.5
	J	4.5	4.5	5.1	5.2	5.1	5.2	5.2	5.1
	K	4.4	4.4	4.8	5.0	4.9	4.8	5.0	5.2
TRK CL	A1	11.1	11.3	9.1	9.0	9.3	9.8	8.5	9.3
	B	12.1	11.5	9.8	9.5	9.8	9.8	9.3	9.5
	C	12.9	12.5	10.3	10.3	10.0	10.0	9.8	9.8
	D	10.1	10.3	8.5	8.8	8.3	8.3	8.5	8.0
	E	11.0	11.4	10.8	10.4	10.4	9.6	9.8	9.2
	A2	11.5	11.4	9.8	9.8	9.6	10.1	9.7	9.3
	F	11.9	11.9	10.2	9.7	9.6	9.8	10.6	9.9
	G	12.1	12.0	10.7	10.4	10.2	9.8	10.1	9.5
	H	13.2	12.7	10.8	10.8	11.1	10.2	10.4	10.9
	I	12.8	12.4	10.5	11.0	11.0	10.4	11.4	10.4
	J	10.5	10.7	9.2	9.4	9.3	9.3	9.4	9.3
	K	11.0	11.3	9.9	9.5	10.2	10.1	9.7	9.4
TRK RMS	A1	40.1	40.2	41.9	42.6	44.4	42.9	41.9	39.9
	B	55.2	53.7	54.4	54.2	53.0	58.7	56.5	57.6
	C	53.0	49.6	46.4	49.5	47.9	45.5	48.8	49.0
	D	45.9	44.2	45.2	47.7	43.9	45.6	45.0	45.6
	E	47.3	45.7	46.3	48.1	48.9	48.3	47.4	46.9
	A2	51.7	50.0	51.8	51.8	51.9	51.7	51.7	50.1
	F	52.5	51.8	49.7	48.1	49.3	51.0	50.4	51.6
	G	50.8	52.2	50.3	47.5	46.8	48.0	46.3	48.0
	H	51.5	51.4	53.9	52.6	51.3	49.9	54.0	52.0
	I	55.0	52.9	51.1	51.1	51.3	50.7	51.1	51.4
	J	50.3	50.5	48.3	47.7	49.2	48.5	46.1	49.9
	K	50.2	52.8	54.0	53.5	54.8	54.4	54.7	54.2

	Group	Train	Start	End	MD-1	MD-2	MD-3	MD-4	MD-5
MTX RT	A1	3358	2851	1860	1716	1735	1746	1663	1371
	B	2866	2295	1515	1447	1197	1268	1134	1098
	C	2357	1942	1337	1364	1328	1185	1092	1010
	D	2268	1967	1413	1591	1374	1234	1218	1228
	E	3575	2585	2412	2244	2256	2072	2394	2182
	A2	2723	2281	1807	1787	1704	1734	1686	1510
	F	1779	1774	1221	1225	1213	1215	1238	1227
	G	2406	1742	1271	1348	1402	1278	1311	1157
	H	1465	1564	1271	1364	1325	1239	1275	1344
	I	1747	1593	1355	1292	1243	1242	1133	1055
	J	1734	1681	1004	1051	1094	1009	1036	1018
	K	1615	1397	976	971	1004	1006	967	920
MTX PC	A1	94.7	88.3	94.5	94.0	91.2	93.4	94.2	91.4
	B	93.9	86.8	92.0	92.0	94.0	90.4	88.5	89.4
	C	95.6	92.2	95.8	90.8	94.8	98.8	96.4	93.9
	D	90.6	86.4	93.3	89.7	86.6	88.6	87.9	87.7
	E	90.9	89.3	91.7	90.4	92.7	92.7	92.0	94.2
	A2	91.9	91.8	93.5	91.6	92.8	94.2	93.0	92.1
	F	92.1	88.1	92.0	90.1	91.7	91.9	92.1	91.8
	G	93.7	88.5	92.8	92.4	95.4	93.2	91.8	93.3
	H	90.7	86.2	91.6	91.1	92.0	91.6	91.1	91.6
	I	89.3	88.9	91.2	89.4	89.0	91.6	87.2	92.4
	J	92.1	89.8	93.8	92.6	92.0	93.5	91.7	92.3
	K	93.2	89.4	93.7	92.2	90.7	93.2	92.3	91.9
MTX TP	A1	29.4	25.8	41.1	43.8	40.4	40.3	43.4	44.0
	B	23.2	24.9	42.2	46.0	54.1	50.2	50.4	53.4
	C	32.4	37.8	49.0	47.5	47.6	56.7	57.4	61.4
	D	34.6	34.2	51.1	46.2	51.8	53.1	51.3	52.0
	E	25.8	29.3	32.0	33.4	34.1	35.6	36.7	40.4
	A2	28.9	31.7	42.1	41.0	41.6	42.7	44.8	43.8
	F	33.7	35.0	48.8	48.0	49.6	50.1	49.6	50.3
	G	29.1	31.5	46.0	44.2	46.0	46.8	46.0	49.0
	H	42.8	36.8	47.0	45.5	49.7	49.4	50.8	50.2
	I	34.6	36.1	43.9	44.2	46.7	47.9	49.0	56.3
	J	36.8	37.4	57.7	54.8	53.1	58.1	56.4	58.6
	K	37.6	38.9	59.6	57.9	55.9	57.9	59.0	61.0

	Group	Train	Start	End	MD-1	MD-2	MD-3	MD-4	MD-5
STN RT	A1	617	682	544	594	560	568	635	572
	B	562	580	485	520	481	452	500	496
	C	598	533	501	503	512	512	516	502
	D	572	580	469	516	482	482	518	515
	E	561	566	525	510	496	510	502	480
	A2	588	592	525	569	533	526	576	577
	F	602	613	557	588	553	567	563	558
	G	571	572	549	562	518	544	552	581
	H	555	577	556	531	591	549	563	515
	I	584	609	544	545	558	569	609	548
	J	590	593	515	530	508	499	557	549
	K	556	587	497	517	532	496	548	522
STN PC	A1	96.9	96.5	98.3	98.6	95.0	98.6	94.8	96.0
	B	96.6	96.6	97.3	95.1	97.5	96.2	95.2	94.9
	C	95.6	96.8	98.7	95.9	98.5	99.0	96.9	96.8
	D	98.0	98.7	97.0	98.4	96.8	97.8	96.4	98.0
	E	98.6	97.8	97.5	98.7	96.9	96.7	96.2	97.0
	A2	98.0	99.0	98.1	98.9	98.5	98.4	97.1	98.3
	F	98.6	98.4	98.4	97.4	98.1	98.8	98.3	98.2
	G	98.4	98.4	98.2	97.9	99.5	98.4	98.0	98.1
	H	98.7	98.7	97.0	98.3	97.7	97.7	96.1	97.1
	I	96.5	98.9	97.8	97.5	97.2	97.7	98.1	97.7
	J	99.0	100.0	98.5	99.1	98.8	99.2	99.1	98.6
	K	98.5	98.9	98.4	98.0	98.5	98.4	98.3	97.7
STN TP	A1	95.2	88.4	109.2	100.6	102.2	105.0	89.4	102.0
	B	104.2	101.3	121.6	111.4	122.3	129.1	115.3	115.8
	C	98.0	111.9	120.6	116.8	118.2	117.7	114.4	117.6
	D	104.2	103.0	125.2	114.3	121.2	122.0	112.4	114.7
	E	108.2	104.5	114.4	117.0	118.3	114.4	116.6	122.4
	A2	103.1	102.7	115.3	108.5	113.9	114.5	104.9	106.3
	F	101.8	99.8	110.6	103.3	109.4	107.3	109.3	107.5
	G	106.0	104.3	109.5	106.0	117.0	112.3	108.3	104.7
	H	109.2	103.7	108.8	114.8	106.5	109.3	105.6	116.0
	I	101.5	100.2	109.5	107.6	107.2	104.3	100.9	108.9
	J	102.8	102.9	116.7	112.9	119.4	120.5	108.8	109.2
	K	109.3	103.7	121.2	115.9	113.5	120.7	110.3	114.0

	Group	Train	Start	End	MD-1	MD-2	MD-3	MD-4	MD-5
CRC RT	A1	1559	1561	1122	1087	1063	1059	1089	962
	B	1422	1267	942	880	826	830	821	803
	C	1436	1204	886	900	837	835	814	796
	D	1432	1390	1041	952	907	932	964	941
	E	1378	1331	1142	1141	1053	1003	997	999
	A2	1254	1208	932	952	929	922	925	878
	F	1452	1447	1144	1081	1103	1066	1040	1047
	G	1298	1259	1015	983	963	908	905	911
	H	1204	1257	849	847	845	833	788	800
	I	1293	1352	965	944	944	982	921	910
	J	1155	1121	887	882	842	856	844	852
	K	1200	1200	981	946	953	950	981	954
CRC PC	A1	95.3	98.1	96.3	97.9	97.9	95.0	96.6	98.2
	B	92.9	92.0	94.6	94.7	91.0	90.6	92.1	90.2
	C	95.3	94.4	95.9	95.7	96.0	97.1	96.4	96.6
	D	86.5	88.7	86.3	85.7	87.3	86.8	86.4	87.2
	E	97.4	97.5	97.3	96.2	96.5	96.1	95.6	96.3
	A2	97.5	97.4	96.9	96.5	97.0	98.2	97.2	96.2
	F	96.9	96.4	97.7	96.2	98.3	98.2	97.4	97.6
	G	97.8	97.7	96.4	97.0	97.1	98.2	98.5	98.0
	H	98.2	97.9	97.8	97.9	97.6	97.4	96.2	97.4
	I	96.4	97.0	97.4	97.0	97.2	97.6	97.0	97.0
	J	97.7	97.2	97.5	98.4	97.3	97.9	98.5	97.1
	K	96.3	97.9	96.0	97.3	96.3	97.0	95.3	92.6
CRC TP	A1	37.8	39.3	53.1	56.6	57.3	53.5	55.3	63.2
	B	39.8	44.3	60.4	64.9	67.0	65.5	67.3	66.4
	C	41.7	48.8	68.7	68.9	73.2	73.6	74.8	76.6
	D	36.6	38.4	53.8	56.6	61.0	60.2	58.4	60.3
	E	45.5	48.3	53.0	52.0	57.1	58.8	60.5	60.6
	A2	48.3	50.2	64.3	62.5	64.6	66.4	64.7	68.2
	F	41.3	41.9	55.5	57.0	55.5	58.2	58.4	60.8
	G	47.5	48.6	59.6	61.8	63.3	67.1	67.1	66.1
	H	50.4	49.6	70.7	69.7	70.9	72.5	74.1	75.5
	I	46.3	44.2	62.0	63.3	63.4	61.5	64.1	65.7
	J	53.5	54.0	67.8	68.8	72.2	70.2	72.2	70.1
	K	48.4	49.6	60.8	64.1	61.8	63.4	61.6	60.7

	Group	Train	Start	End	MD-1	MD-2	MD-3	MD-4	MD-5
MAN RT	A1	2245	2371	1686	1743	1676	1739	1544	1676
	B	2451	2357	1552	1558	1415	1460	1407	1449
	C	1895	1800	1388	1377	1426	1370	1285	1346
	D	2078	2208	1495	1472	1383	1405	1459	1371
	E	2049	2047	1820	1722	1692	1603	1501	1451
	A2	1931	1892	1398	1483	1454	1425	1390	1381
	F	2055	2094	1652	1683	1598	1626	1605	1571
	G	1868	1762	1496	1555	1458	1462	1446	1399
	H	1923	2042	1433	1530	1493	1393	1409	1339
	I	2046	2192	1553	1523	1540	1514	1496	1431
	J	2168	2169	1521	1563	1523	1545	1515	1453
	K	1856	1835	1347	1400	1400	1365	1377	1351
MAN PC	A1	94.6	90.2	97.0	97.2	97.1	96.5	98.4	98.4
	B	89.6	90.4	95.6	94.3	96.5	95.8	95.3	95.8
	C	96.3	97.9	98.1	96.2	96.9	97.7	97.3	97.9
	D	91.0	88.9	96.7	98.2	97.8	94.5	96.8	97.3
	E	98.0	98.9	98.3	99.2	98.5	99.2	98.0	98.5
	A2	95.2	94.3	97.9	97.6	97.5	98.3	97.5	98.2
	F	95.4	94.3	98.4	98.0	98.5	98.0	98.3	98.5
	G	96.6	95.2	98.5	97.6	99.2	97.9	97.5	98.5
	H	90.8	90.6	96.4	96.2	97.9	97.7	97.5	98.0
	I	95.2	94.9	97.4	97.5	96.8	98.0	97.0	96.9
	J	92.6	93.9	98.9	99.3	99.8	98.5	98.8	99.2
	K	93.5	93.6	98.7	99.1	99.2	98.9	98.6	98.5
MAN RTX	A1	-	-	-	-	-	-	-	-
	B	-	-	-	-	-	-	-	-
	C	-	-	-	-	-	-	-	-
	D	-	-	-	-	-	-	-	-
	E	-	-	-	-	-	-	-	-
	A2	-	1957	1522	1588	1583	1521	1489	1522
	F	-	2237	-	1821	1779	1874	1833	1776
	G	-	1892	1618	1668	1572	1547	1573	1568
	H	2051	2201	1631	1758	1789	1639	1683	1571
	I	2140	2244	1729	1741	1766	1701	1737	1589
	J	2160	2199	1640	1628	1618	1654	1616	1575
	K	1858	1915	1446	1520	1554	1458	1518	1492
MAN PCX	A1	-	-	-	-	-	-	-	-
	B	-	-	-	-	-	-	-	-
	C	-	-	-	-	-	-	-	-
	D	-	-	-	-	-	-	-	-
	E	-	-	-	-	-	-	-	-
	A2	-	94.1	97.8	97.1	96.5	98.0	97.4	97.8
	F	-	92.5	-	98.7	99.2	96.3	95.9	97.8
	G	-	97.3	98.9	96.7	98.6	96.8	97.0	99.3
	H	89.8	88.8	94.6	95.0	97.3	97.7	97.8	98.4
	I	95.4	94.4	96.7	96.7	97.8	97.6	96.8	95.2
	J	91.2	92.7	98.6	98.9	99.7	97.4	98.9	98.5
	K	92.8	90.8	99.1	99.1	99.6	98.1	96.6	98.8

	Group	Train	Start	End	MD-1	MD-2	MD-3	MD-4	MD-5
MTH RT	A1	2299	2482	2051	2067	2029	2234	2104	1989
	B	2315	2392	1673	1609	1538	1495	1518	1438
	C	1902	1787	1532	1537	1566	1461	1438	1400
	D	2359	2463	2214	2221	2143	2240	2139	2046
	E	2210	2288	2107	2048	1995	1889	1886	1897
	A2	2111	2160	1753	1858	1856	1717	1779	1672
	F	2378	2435	2001	2052	2030	1892	1925	1863
	G	2266	2194	1930	1890	1882	1810	1881	1815
	H	2198	2326	1824	1937	1853	1819	1734	1781
	I	2251	2334	1944	2063	1940	1938	1895	1781
	J	2183	2269	1832	1892	1915	1735	1816	1771
	K	2118	2132	1690	1767	1843	1672	1779	1624
MTH PC	A1	94.5	93.1	93.2	94.6	90.1	91.6	89.6	94.6
	B	93.0	92.2	94.0	94.6	92.4	94.0	94.7	94.3
	C	94.5	94.2	98.3	94.4	97.0	97.6	94.9	96.6
	D	93.7	89.7	96.8	95.0	91.6	93.0	93.3	96.5
	E	95.4	95.9	97.9	98.3	96.8	98.1	97.8	95.1
	A2	94.0	92.8	96.4	95.6	96.5	97.9	96.3	96.3
	F	91.8	88.0	97.0	95.4	96.5	94.6	96.1	96.1
	G	92.8	86.4	95.9	95.6	97.3	97.5	96.7	98.3
	H	91.7	90.8	95.8	95.7	97.2	95.8	95.3	96.2
	I	93.4	87.0	96.0	97.0	96.2	94.4	96.1	96.1
	J	93.1	93.1	97.9	95.8	97.0	98.0	97.7	97.7
	K	95.1	92.9	98.2	97.7	98.5	97.9	98.4	96.8
MTH RTX	A1	-	-	-	-	-	-	-	-
	B	-	-	-	-	-	-	-	-
	C	-	-	-	-	-	-	-	-
	D	-	-	-	-	-	-	-	-
	E	-	-	-	-	-	-	-	-
	A2	-	2274	1993	2105	2035	1919	2013	1891
	F	-	2607	-	2242	2190	2034	2114	2015
	G	-	2159	2135	2080	1975	1954	2032	1982
	H	2457	2440	2080	2211	2156	2066	2019	2043
	I	2456	2482	2143	2261	2098	2113	2018	1962
	J	2301	2318	2025	2116	2064	1865	1996	1949
	K	2281	2197	1930	2039	2083	1871	2050	1856
MTH PCX	A1	-	-	-	-	-	-	-	-
	B	-	-	-	-	-	-	-	-
	C	-	-	-	-	-	-	-	-
	D	-	-	-	-	-	-	-	-
	E	-	-	-	-	-	-	-	-
	A2	-	92.6	97.2	95.3	95.3	98.0	96.4	95.5
	F	-	88.7	-	95.6	95.6	95.8	95.8	94.7
	G	-	87.9	95.1	94.1	97.8	96.9	97.4	98.2
	H	88.4	89.4	94.0	96.9	98.2	97.2	95.3	97.2
	I	91.0	86.9	95.5	95.9	94.9	96.8	96.8	96.9
	J	90.1	93.6	97.5	97.2	96.4	98.5	97.6	97.4
	K	94.1	93.0	98.5	98.2	98.5	97.5	98.7	98.5

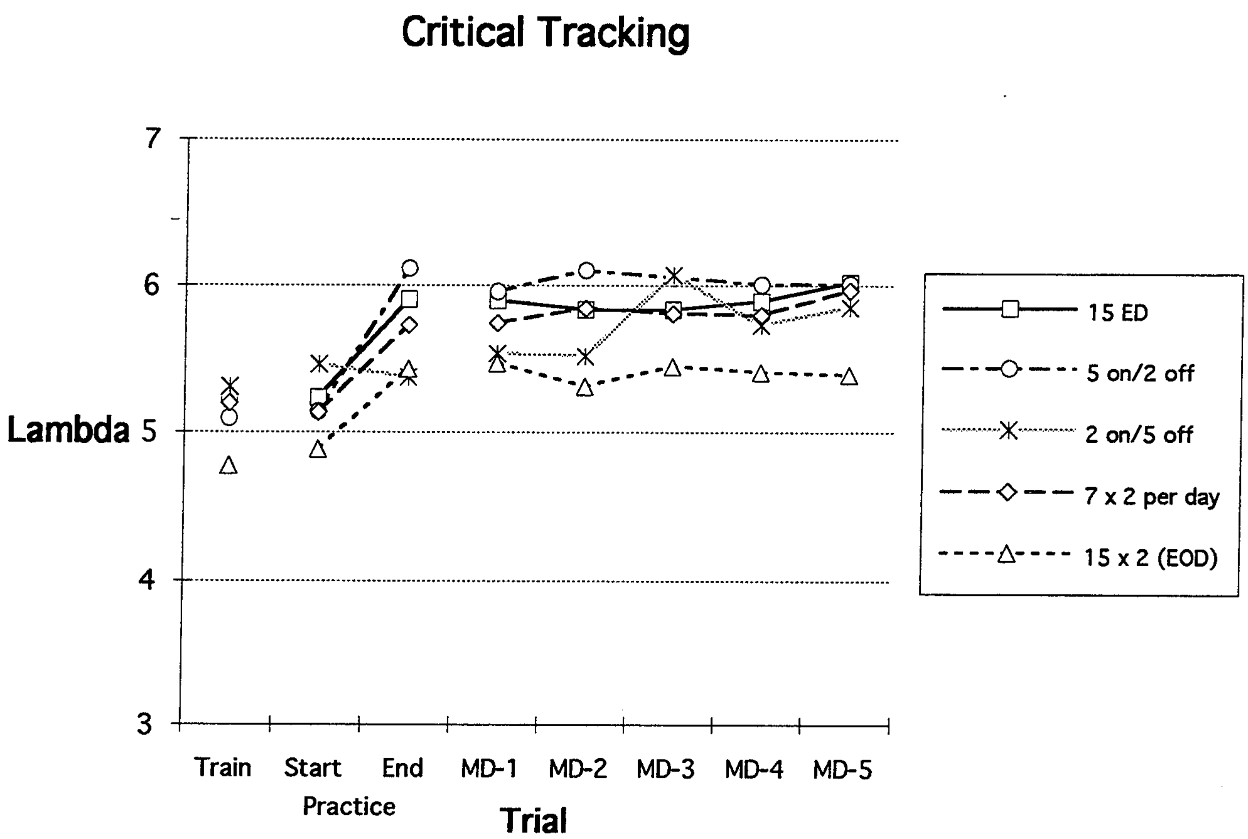
	Group	Train	Start	End	MD-1	MD-2	MD-3	MD-4	MD-5
DUL CL	A1	0.3	0.0	7.5	6.8	5.0	11.3	5.0	6.3
	B	2.0	1.0	7.3	15.5	4.8	7.8	6.3	6.5
	C	10.1	11.0	11.6	10.8	13.3	12.5	8.0	7.0
	D	0.0	0.0	3.5	5.0	4.3	2.5	4.3	3.8
	E	2.9	2.6	24.6	16.4	19.0	11.8	15.2	9.4
	A2	1.5	2.1	14.4	14.3	9.5	10.5	11.7	12.9
	F	2.3	2.3	13.0	10.5	16.0	9.1	10.2	10.2
	G	1.3	2.0	14.3	13.9	9.8	10.2	8.4	7.7
	H	2.6	2.2	9.2	12.1	7.8	6.7	8.8	8.9
	I	2.5	1.6	13.6	11.4	13.2	14.4	9.4	14.5
	J	0.4	0.2	4.8	6.8	4.0	6.1	5.1	4.9
	K	1.7	1.8	11.4	7.2	10.0	11.0	12.3	8.3
DUL RMS	A1	15.3	9.1	41.1	37.4	43.8	44.7	41.0	36.4
	B	43.4	28.4	56.3	60.5	54.1	55.3	55.8	55.5
	C	42.9	44.4	44.4	51.5	44.4	47.7	48.7	44.2
	D	19.1	13.5	39.7	46.3	44.4	38.1	44.3	41.9
	E	35.2	33.0	54.7	55.9	55.9	57.2	58.3	51.6
	A2	32.7	31.7	55.4	57.4	56.3	56.0	49.4	50.4
	F	38.4	37.6	54.4	54.8	52.8	55.4	55.6	57.2
	G	30.2	32.0	50.5	51.6	43.8	46.5	43.2	43.3
	H	36.0	34.1	50.9	52.8	49.5	52.4	54.9	49.8
	I	40.7	37.4	56.0	53.4	50.2	52.4	53.5	54.6
	J	26.6	26.7	45.9	47.1	41.0	43.6	44.6	40.0
	K	28.7	29.2	52.5	50.0	47.2	51.4	51.4	50.1

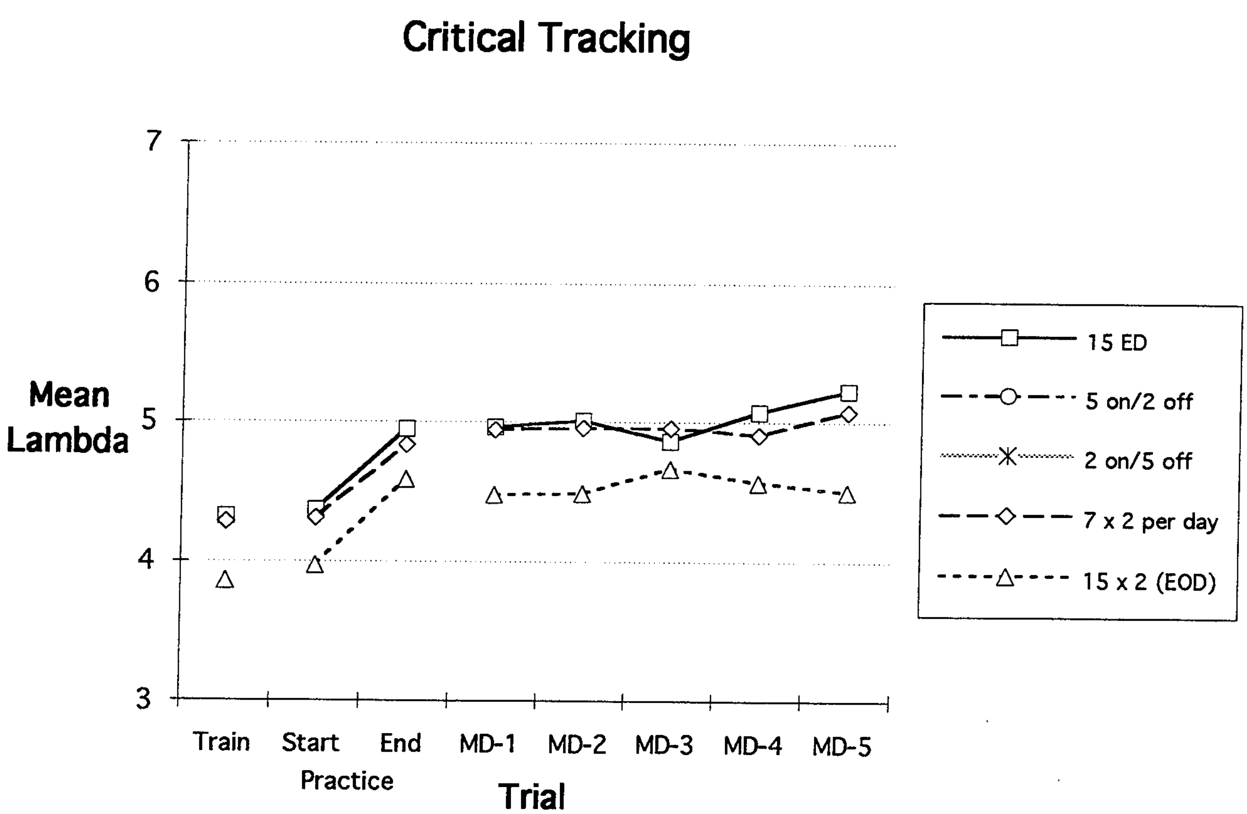
	Group	Train	Start	End	MD-1	MD-2	MD-3	MD-4	MD-5
DUL RT	A1	735	771	847	859	849	990	865	892
	B	700	664	652	644	566	608	594	617
	C	678	629	652	626	700	643	576	626
	D	636	633	601	603	605	589	600	617
	E	653	644	675	642	686	618	694	646
	A2	705	687	712	694	655	680	651	725
	F	703	665	703	707	718	723	679	708
	G	654	697	673	655	612	660	637	664
	H	679	698	692	659	688	652	690	656
	I	704	685	723	671	685	701	752	720
	J	689	719	703	682	648	715	671	699
	K	654	704	697	639	632	688	643	678
DUL PC	A1	97.1	94.2	94.9	96.7	97.8	94.6	94.5	93.4
	B	96.3	96.3	96.9	96.7	96.8	97.1	96.1	96.0
	C	97.4	98.2	98.2	96.0	98.0	98.1	98.6	94.5
	D	97.2	96.7	96.1	96.4	95.5	97.1	98.2	94.9
	E	97.9	97.3	96.1	96.7	98.6	98.0	96.4	97.1
	A2	96.6	98.3	97.1	98.6	98.0	97.6	98.1	96.0
	F	97.2	97.0	97.5	97.3	97.5	97.7	97.7	97.4
	G	98.2	98.2	97.4	99.0	98.5	98.8	98.4	97.9
	H	98.3	96.9	97.4	98.5	97.0	98.9	97.3	96.3
	I	97.3	97.7	97.7	98.4	95.5	98.0	97.7	97.3
	J	98.1	98.3	98.4	96.8	98.7	97.7	98.6	97.2
	K	98.3	98.6	97.0	98.3	98.3	97.2	97.9	97.3
DUL TP	A1	80.1	73.3	71.2	74.6	75.2	66.1	73.2	74.0
	B	86.3	90.2	92.3	93.6	105.6	100.6	100.0	94.8
	C	88.4	97.4	93.9	96.1	92.8	95.4	105.0	93.7
	D	91.9	91.2	97.3	96.7	96.5	100.6	99.3	92.9
	E	92.7	92.6	86.2	91.4	88.2	96.1	84.9	91.3
	A2	88.1	88.9	87.1	90.5	94.8	91.2	94.3	84.7
	F	86.1	89.5	86.1	85.5	86.8	83.8	89.0	84.8
	G	91.7	86.6	88.8	93.1	98.3	92.8	95.6	89.5
	H	90.8	87.1	88.7	92.8	90.2	95.2	88.5	91.5
	I	85.7	88.2	83.3	89.8	85.6	87.0	82.4	84.7
	J	88.2	84.4	88.6	88.1	94.6	86.8	94.3	86.1
	K	94.4	88.9	92.5	97.1	98.6	92.7	96.1	91.0

	Group	Train	Start	End	MD-1	MD-2	MD-3	MD-4	MD-5
FAT RP	A1	4.1	2.5	2.5	3.0	2.8	4.3	2.3	3.0
	B	3.5	3.3	2.3	3.0	2.5	2.3	2.0	2.0
	C	4.0	3.5	2.5	3.3	2.8	2.8	3.0	2.5
	D	3.4	2.3	2.6	3.5	2.8	3.8	3.8	4.3
	E	3.4	3.2	2.7	3.0	3.2	3.0	3.2	2.8
	A2	3.6	3.4	3.1	3.2	3.3	3.1	3.4	3.0
	F	3.6	2.9	3.7	3.8	3.5	3.8	2.8	3.2
	G	4.1	3.0	3.9	4.0	3.6	3.2	3.5	3.1
	H	3.1	3.4	3.3	3.4	3.8	2.9	3.6	3.1
	I	2.8	2.9	2.2	3.1	3.3	2.4	2.6	2.0
	J	3.8	3.4	4.0	3.8	3.7	4.1	4.3	3.4
	K	3.6	3.6	3.3	2.8	3.5	3.1	3.4	2.6
FAT RT	A1	-	-	-	-	-	-	-	-
	B	-	-	-	-	-	-	-	-
	C	-	-	-	-	-	-	-	-
	D	-	-	-	-	-	-	-	-
	E	-	-	-	-	-	-	-	-
	A2	-	-	-	-	-	-	-	-
	F	-	-	-	-	-	-	-	-
	G	-	-	-	-	-	-	-	-
	H	3.2	4.5	2.2	1.9	2.1	2.8	1.6	2.1
	I	5.0	4.8	2.7	2.6	2.1	2.3	2.4	2.1
	J	4.6	4.6	2.7	3.8	2.2	3.1	2.1	2.9
	K	4.3	6.1	3.1	4.2	2.2	4.3	3.1	2.7

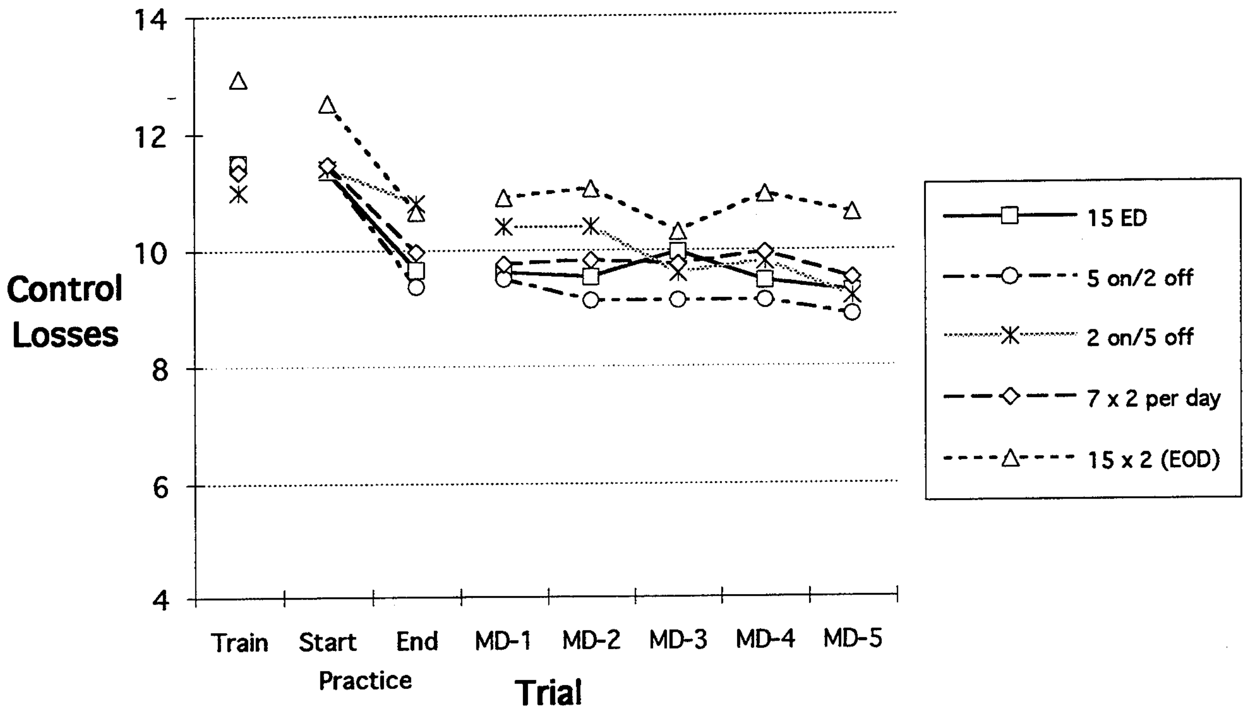
APPENDIX F

PRACTICE SCHEDULE DATA

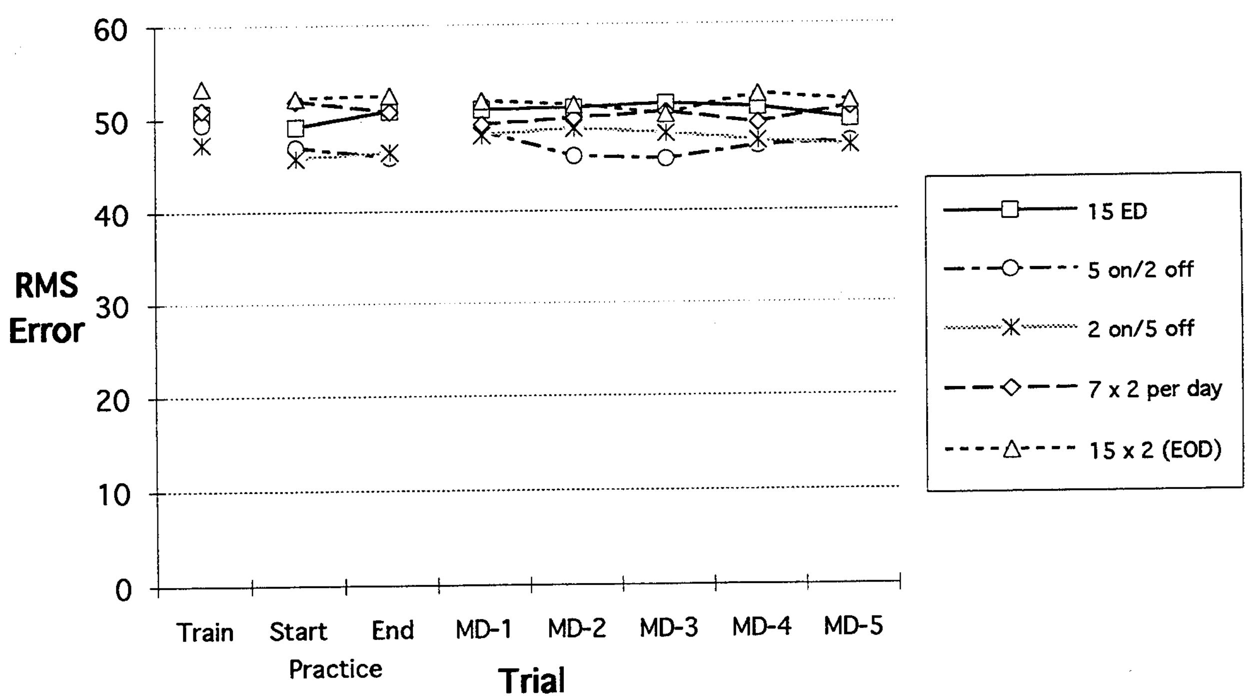

Critical Tracking
7
6
Lambda 5
4
3
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

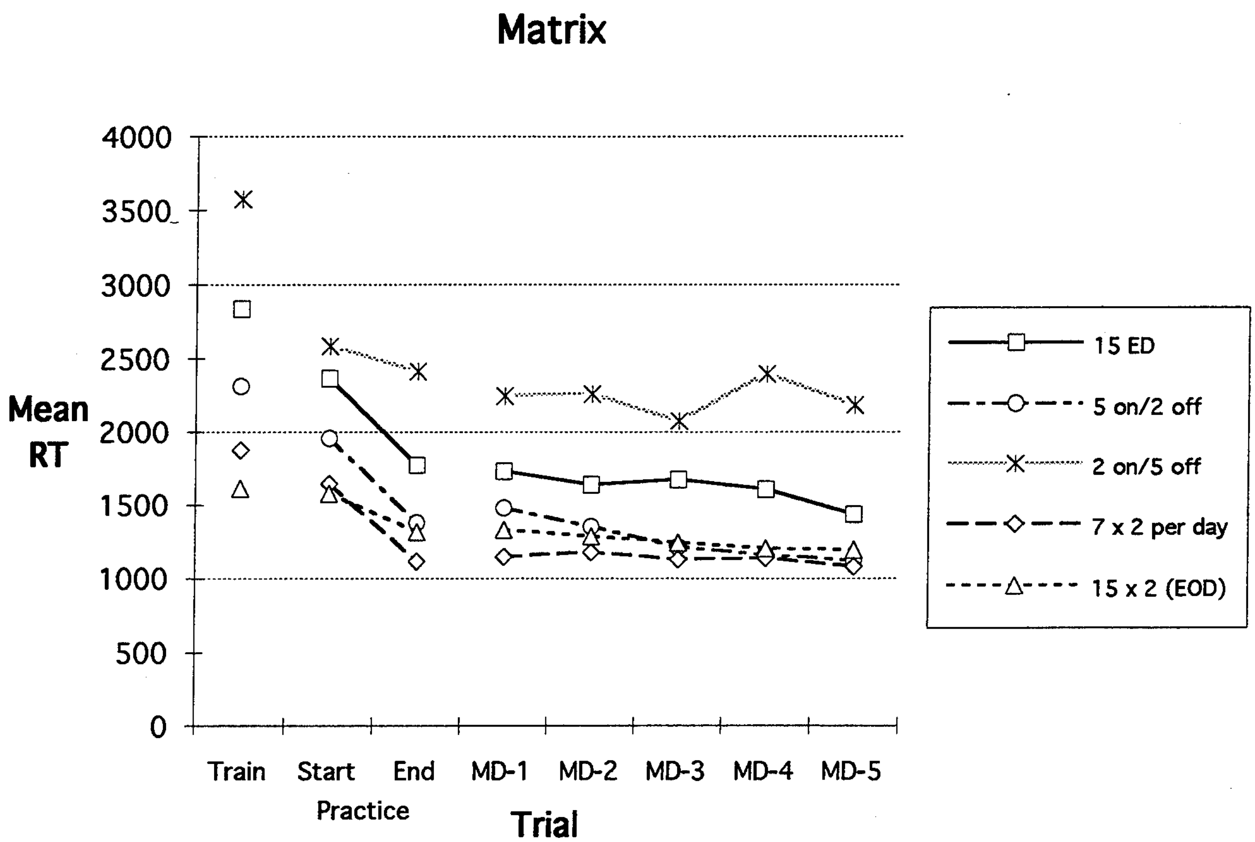

Critical Tracking
7
6
Mean Lambda
5
4
3
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

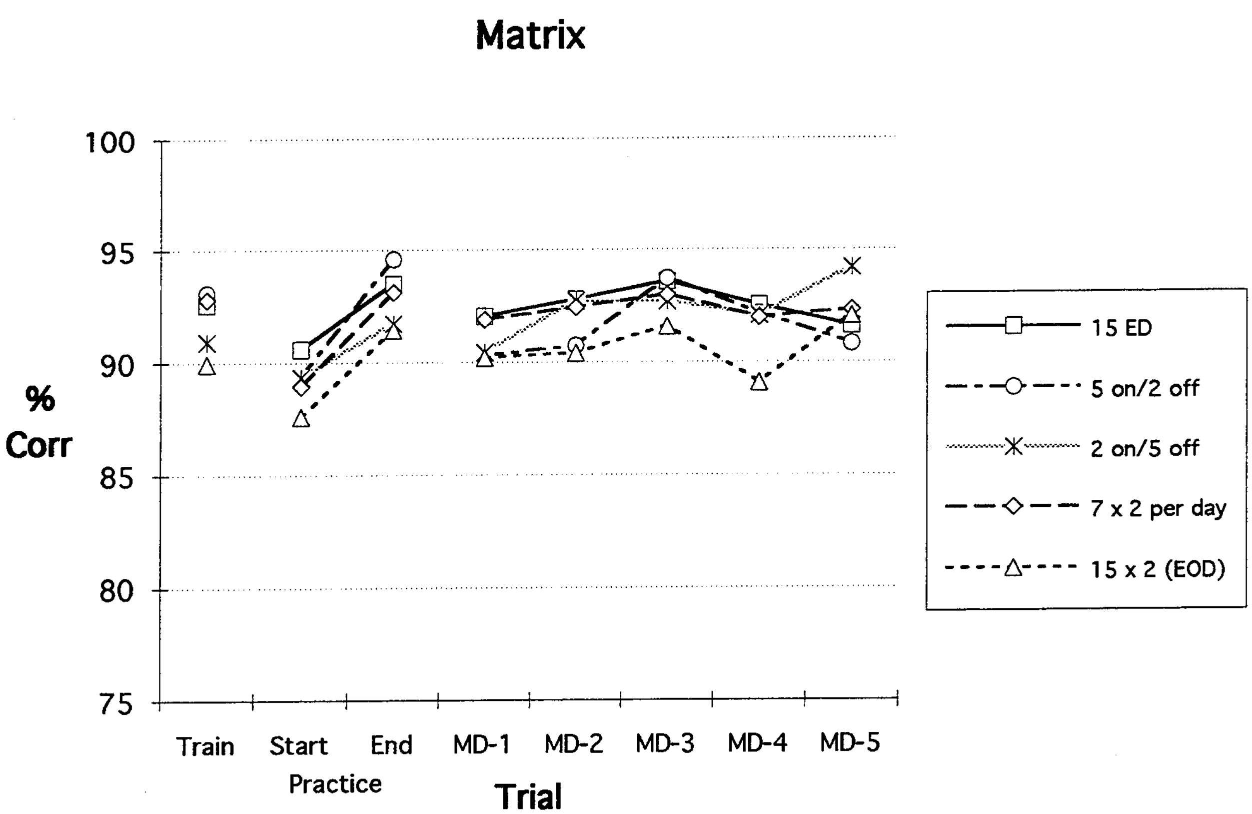
Critical Tracking

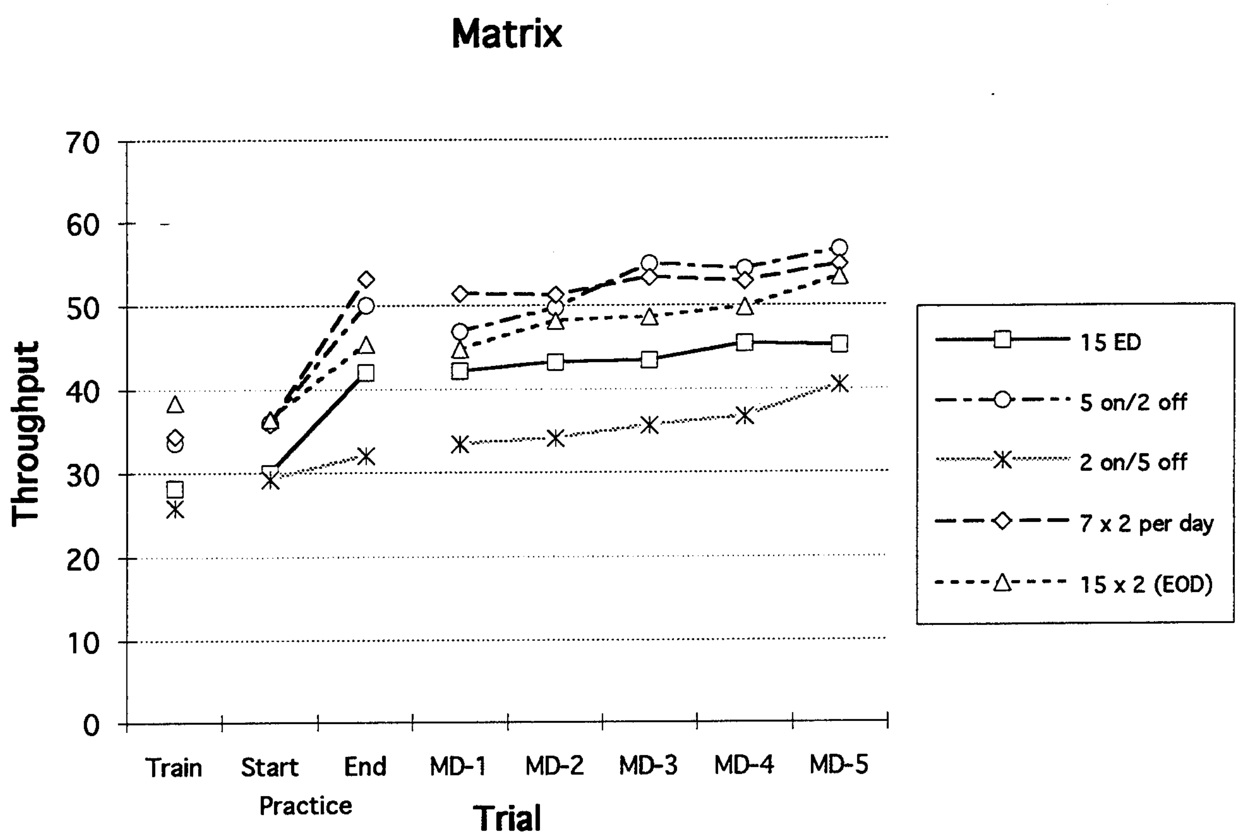


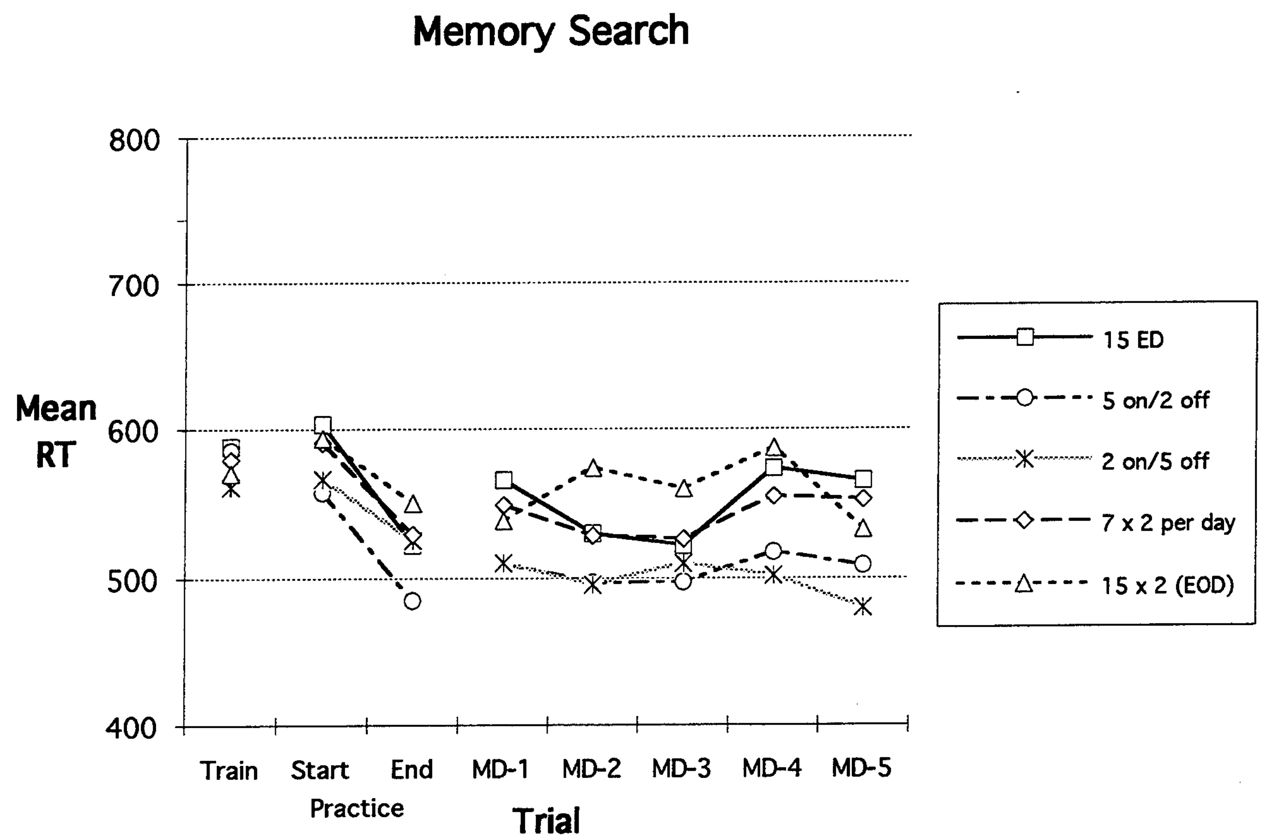
Critical Tracking

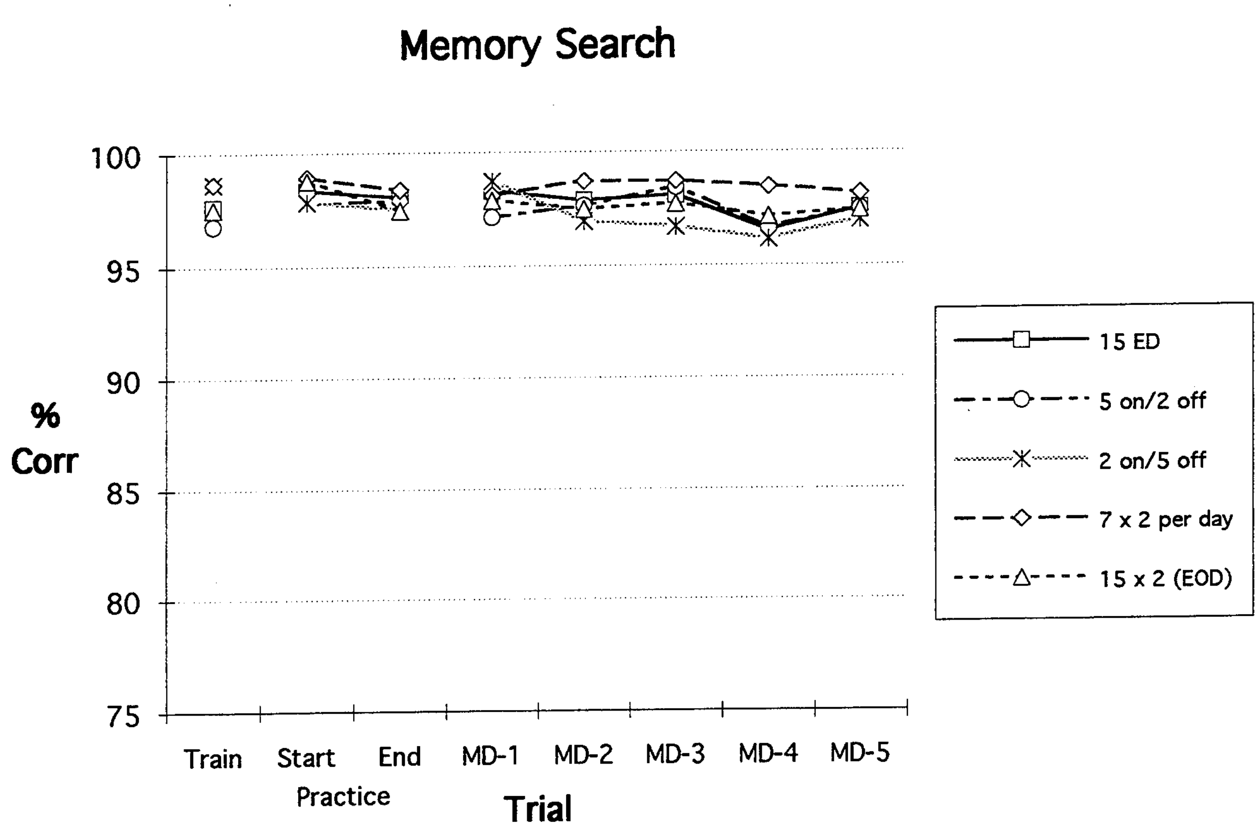


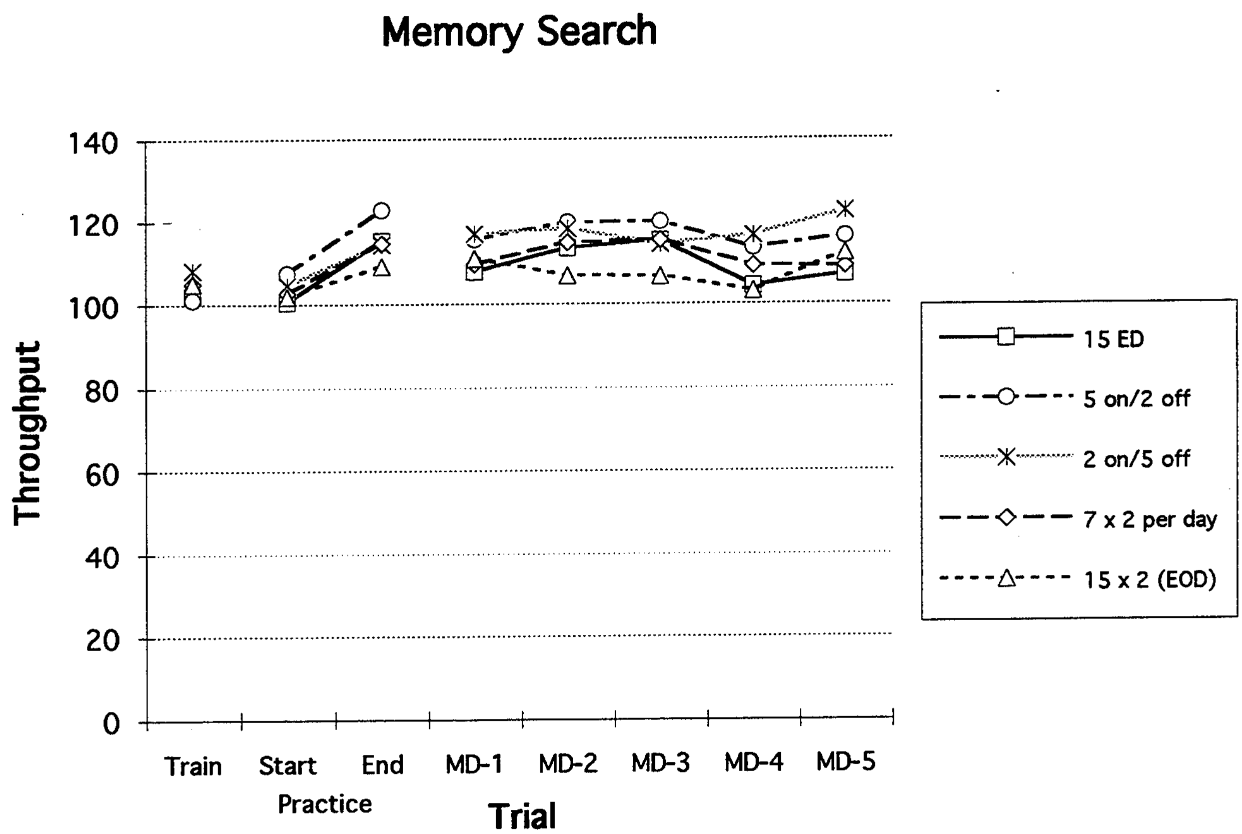

Matrix
4000
3500
3000
2500
2000
1500
1000
500
0
Mean RT
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

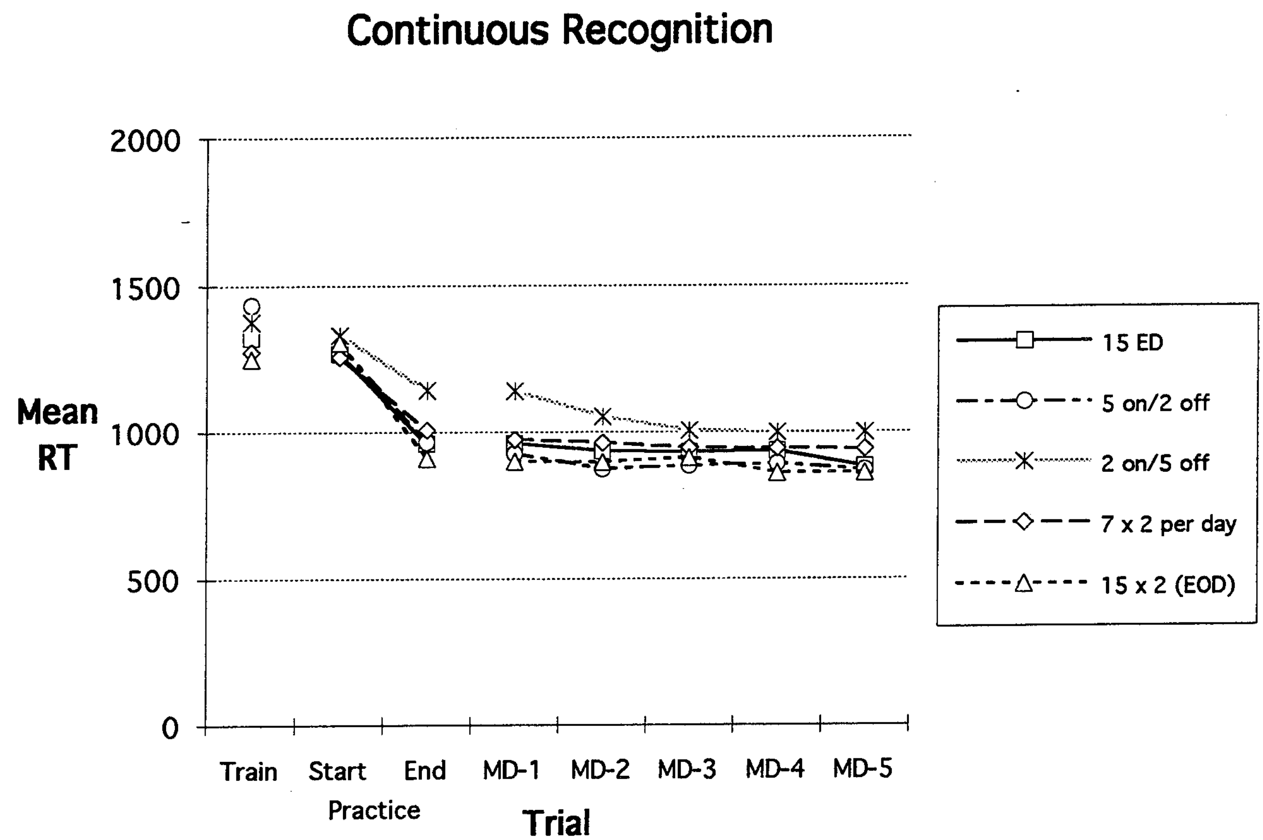

Matrix
100
95
90
85
80
75
% Corr
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

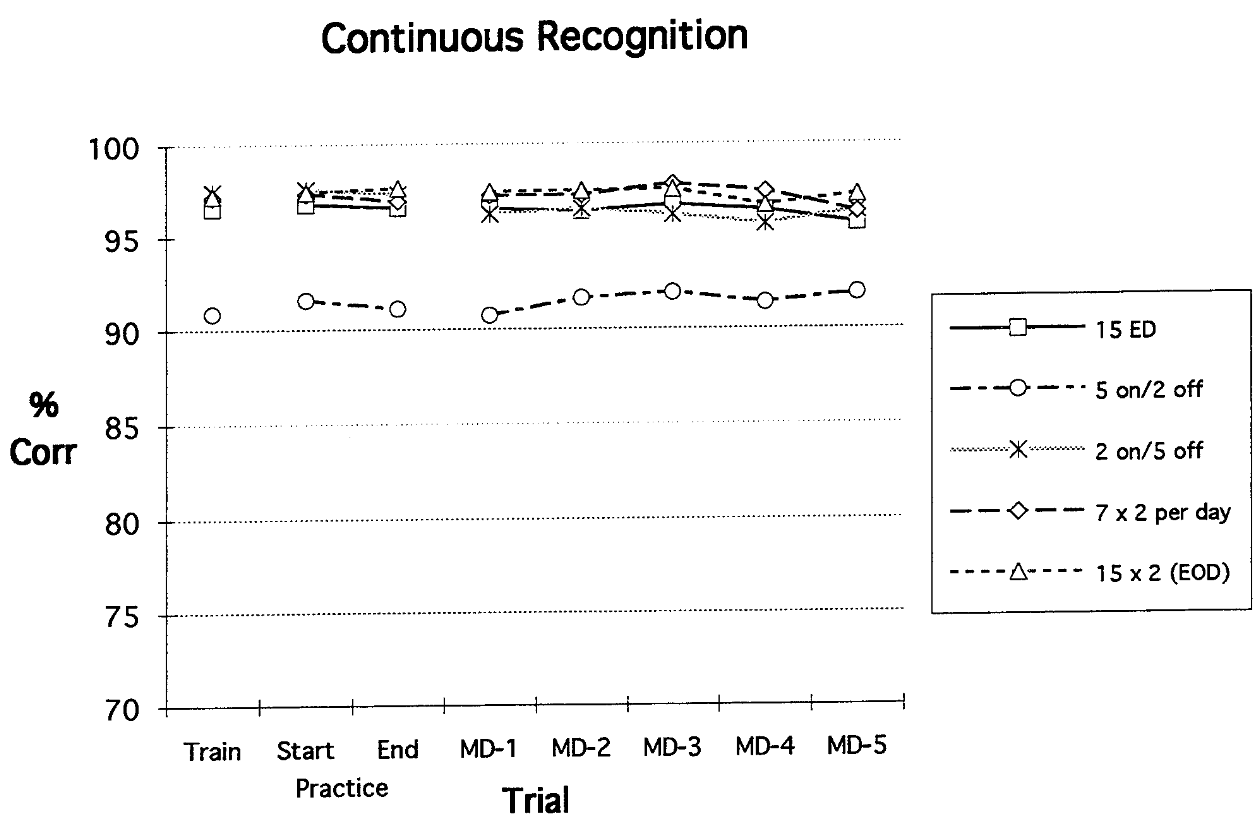

Matrix
70
60
50
40
30
20
10
0
Throughput
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

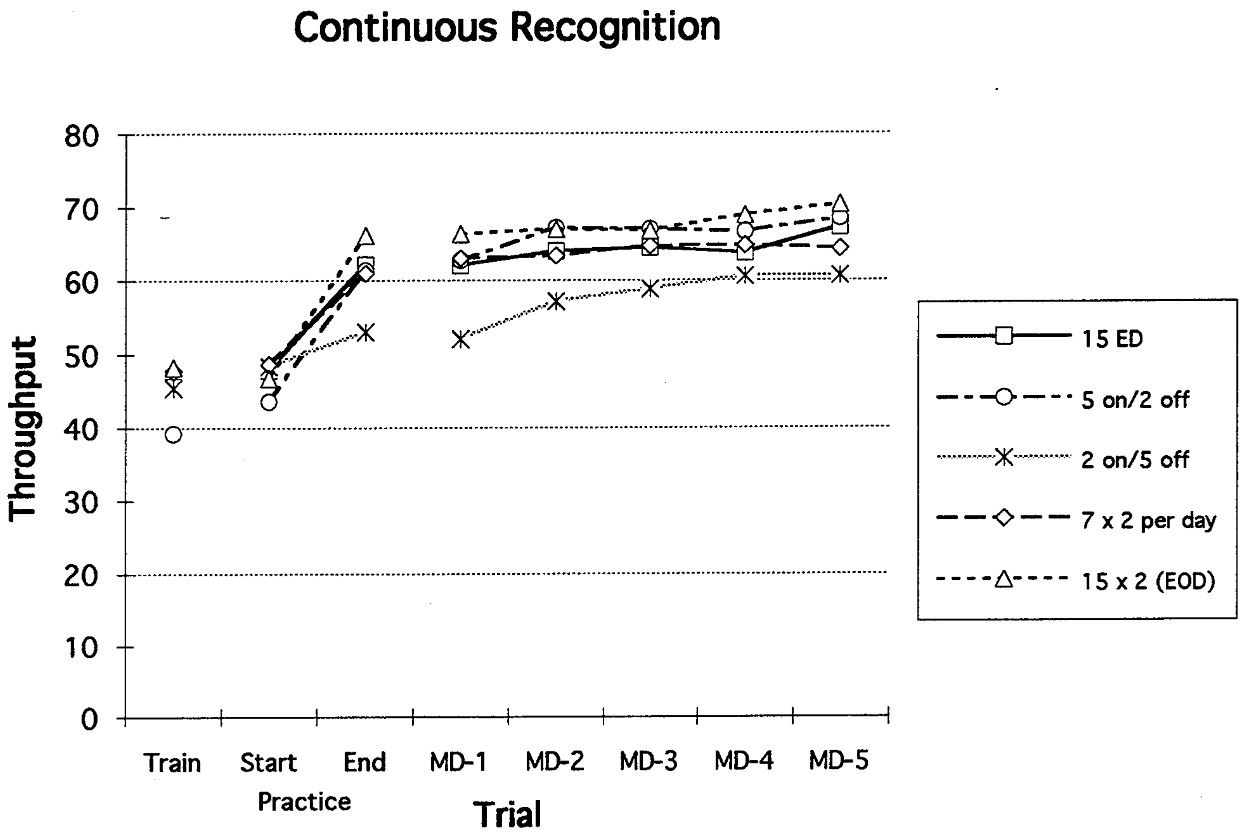

Memory Search
Mean RT
800
700
600
500
400
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)


Memory Search
% Corr
100
95
90
85
80
75
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

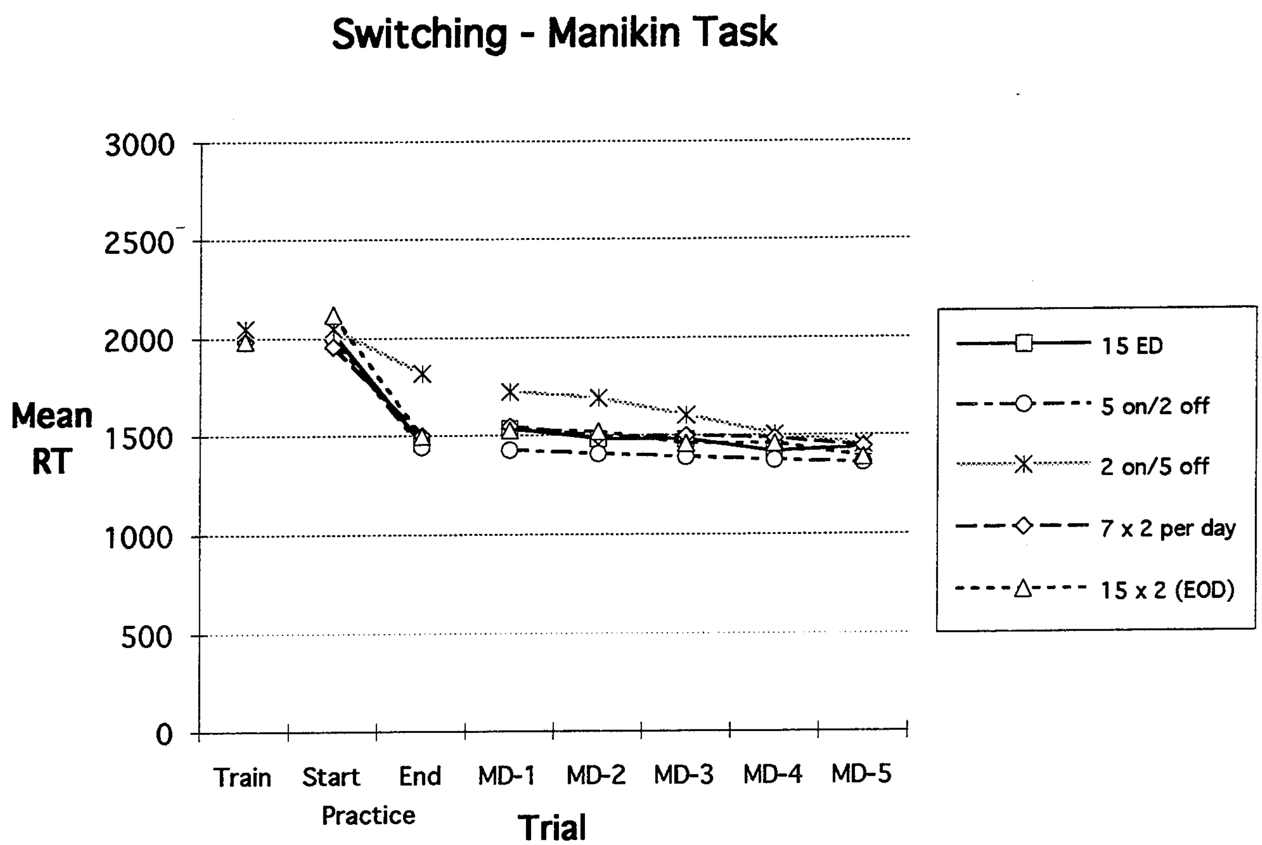

Memory Search
Throughput
140
120
100
80
60
40
20
0
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

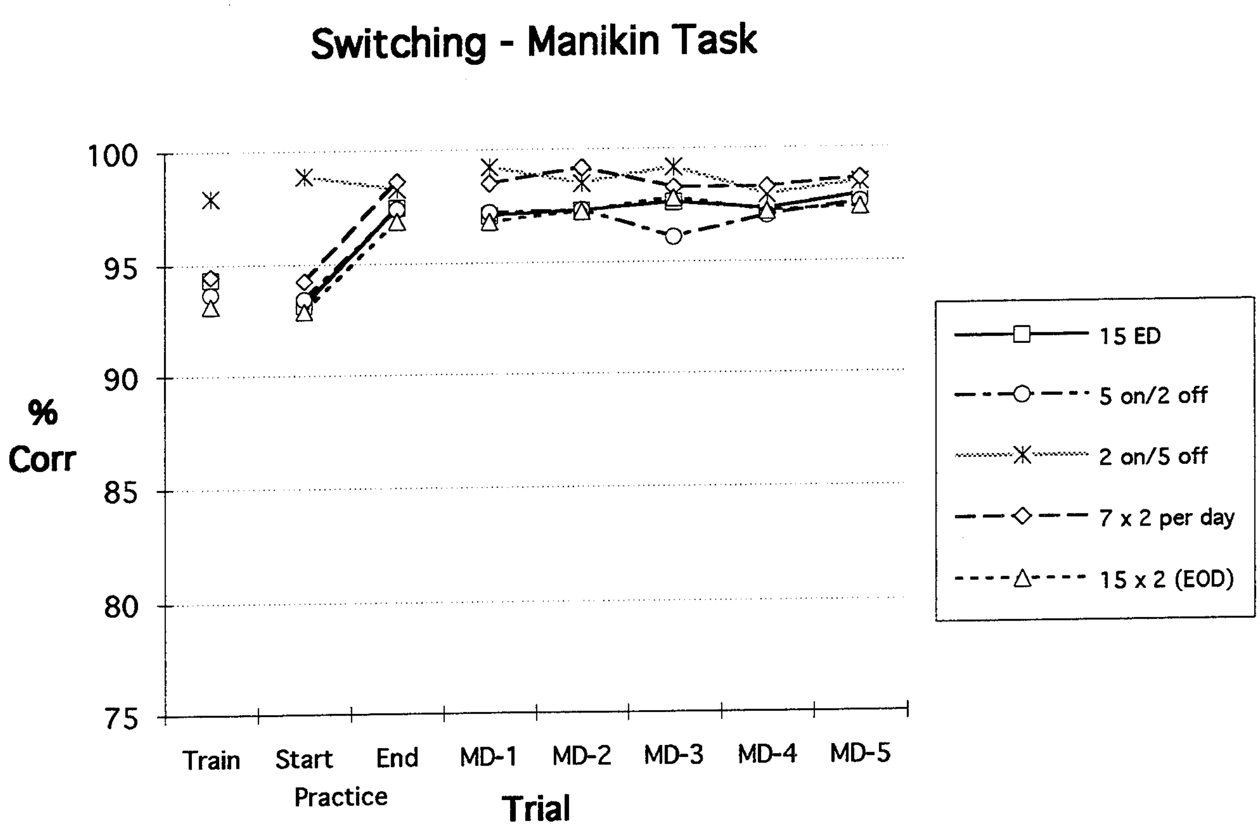

Continuous Recognition
2000
1500
1000
500
0
Mean RT
Train
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Practice
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

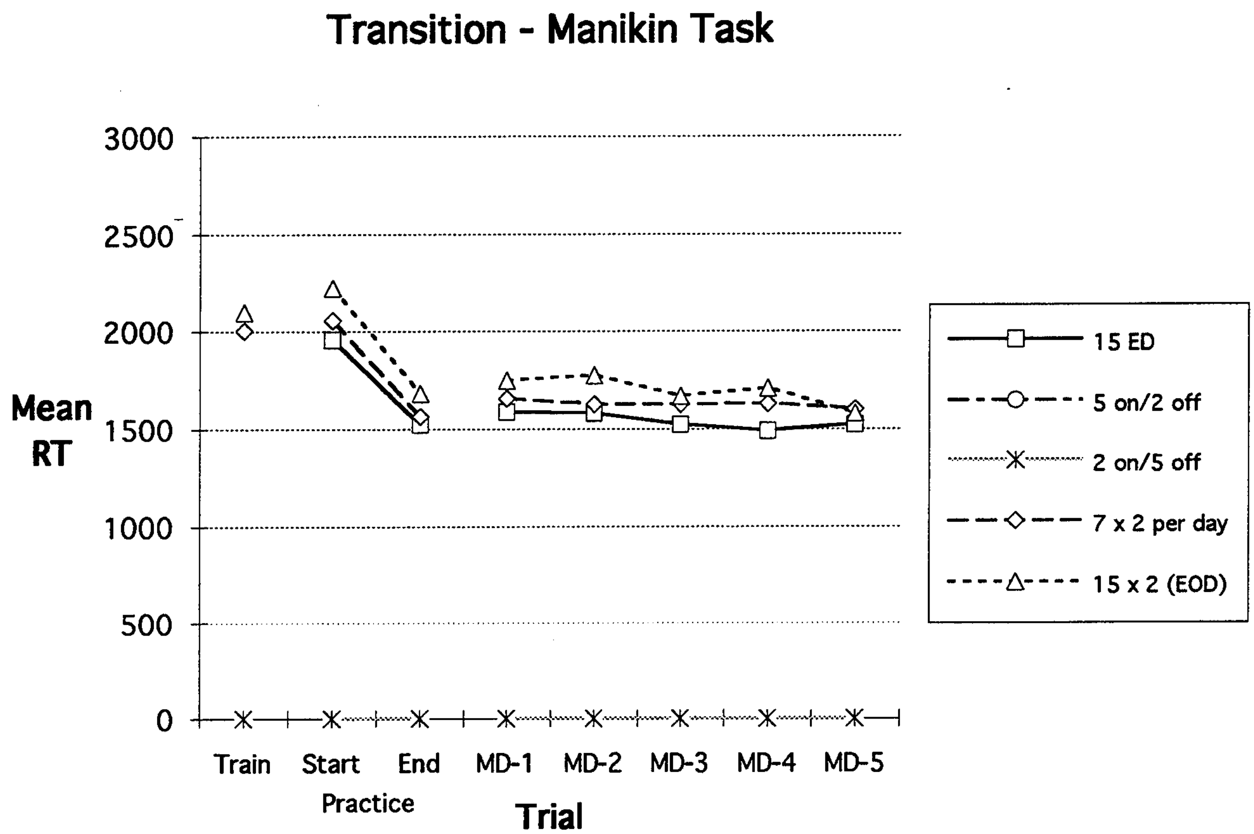

Continuous Recognition
100
95
90
85
80
75
70
% Corr
Train
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Practice
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)



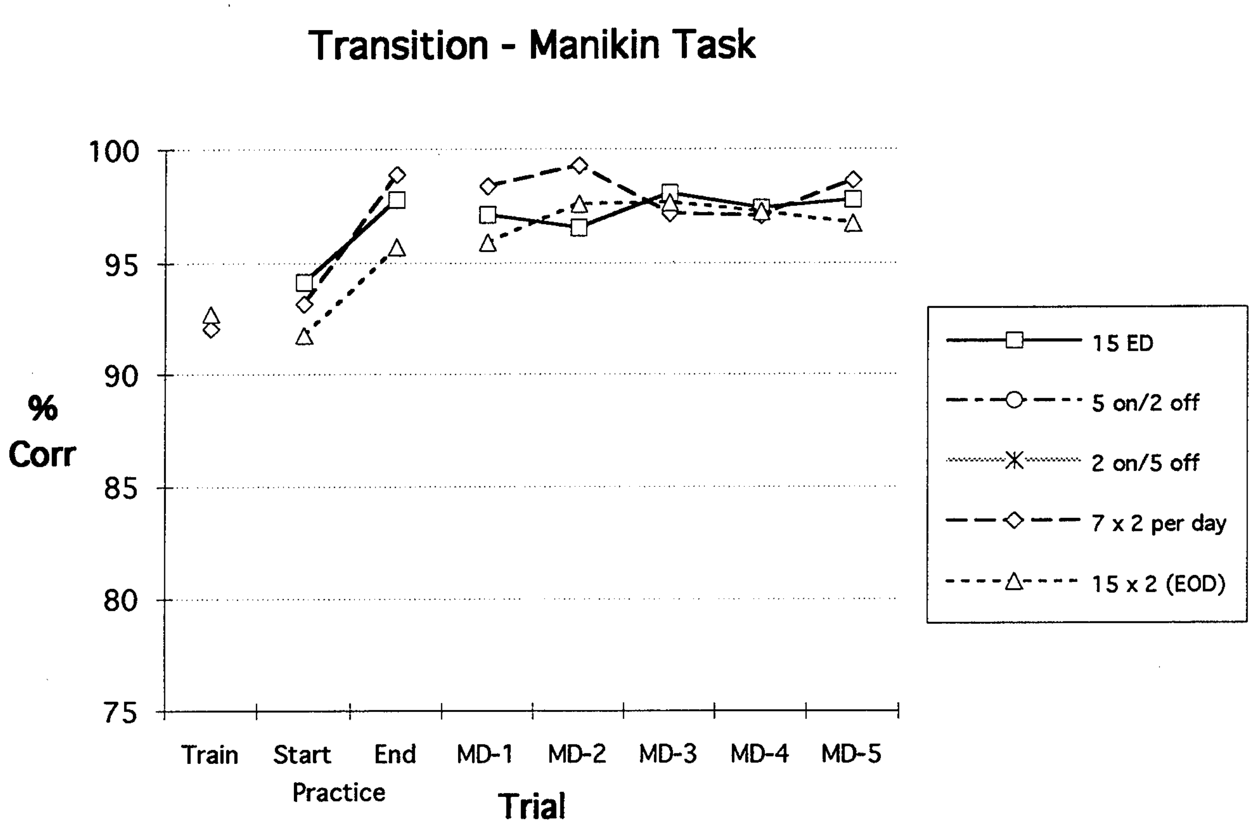
Continuous Recognition
Throughput
80
70
60
50
40
30
20
10
0
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)


Switching - Manikin Task
3000
2500
2000
1500
1000
500
0
Mean RT
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

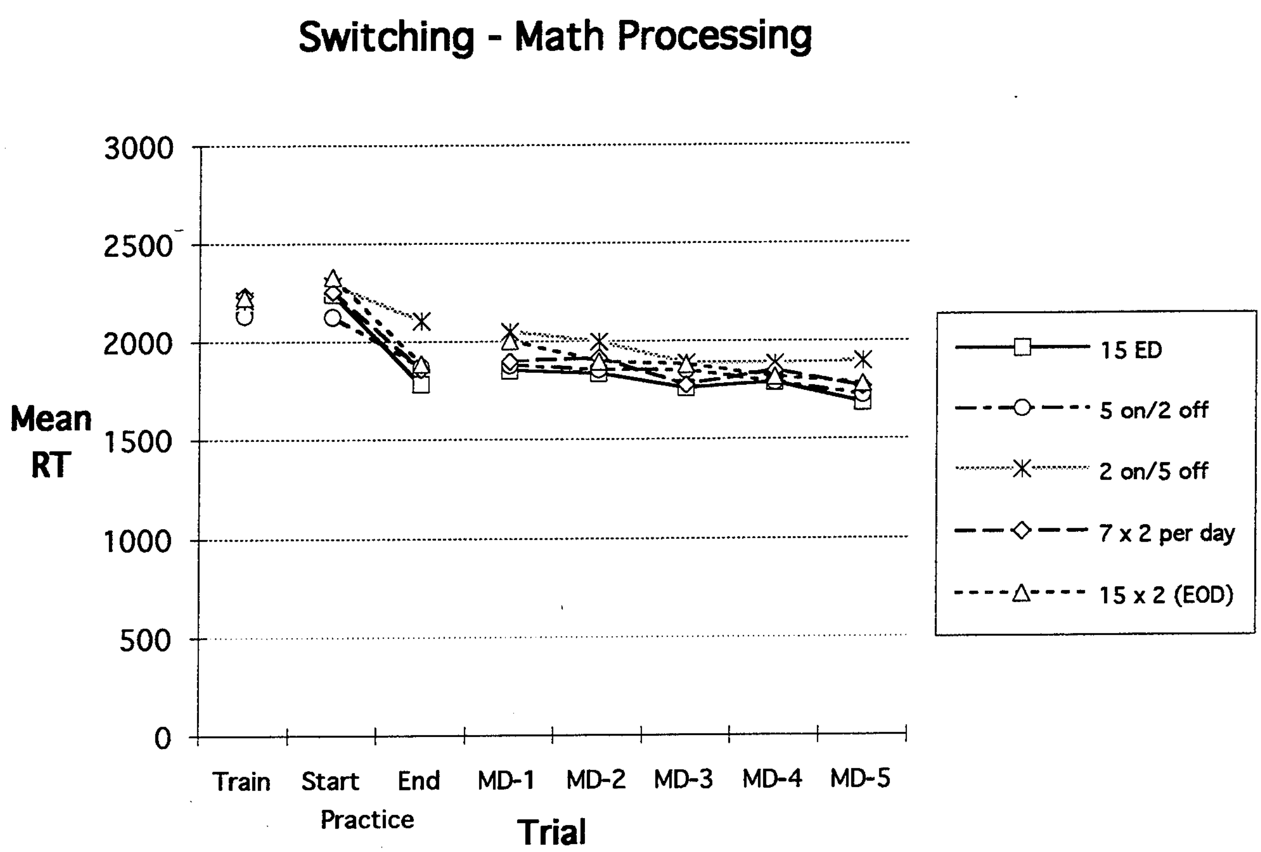

Switching - Manikin Task
100
95
90
85
80
75
% Corr
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

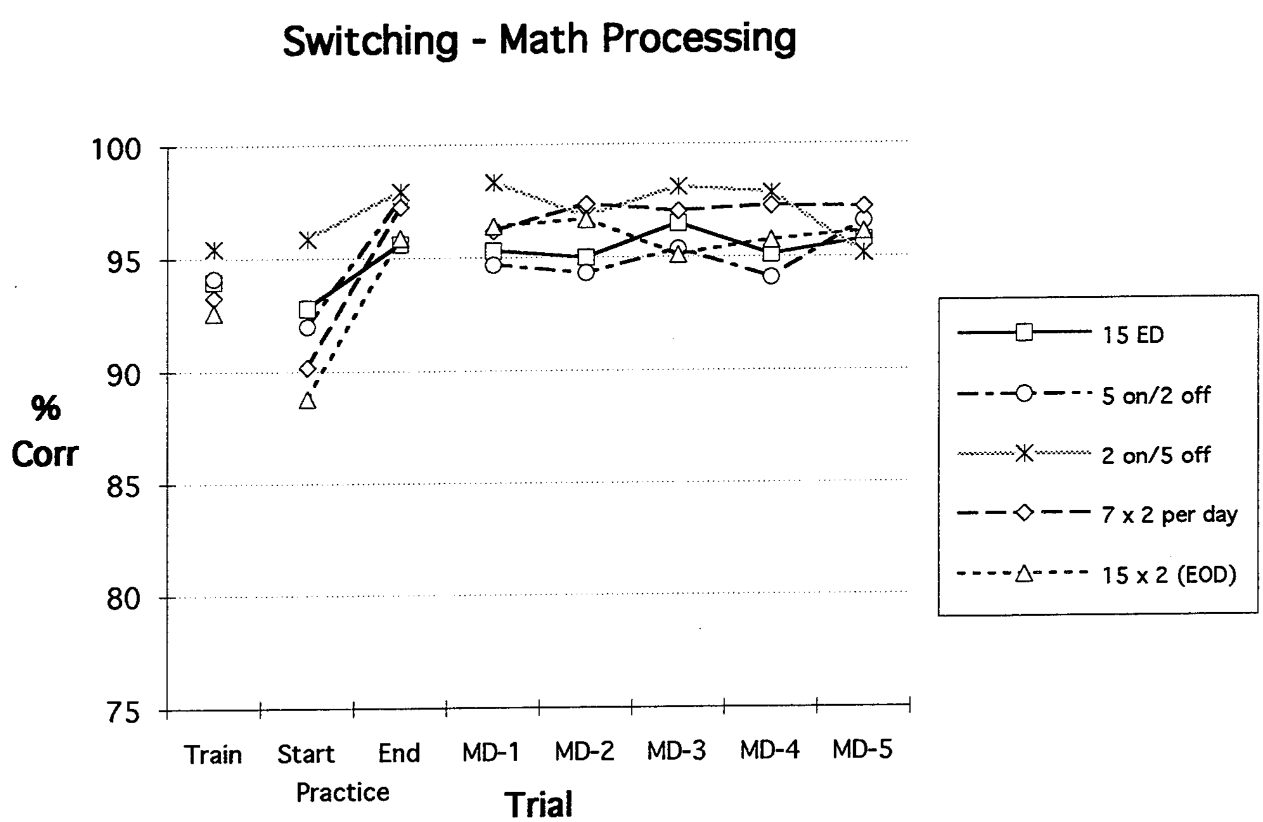


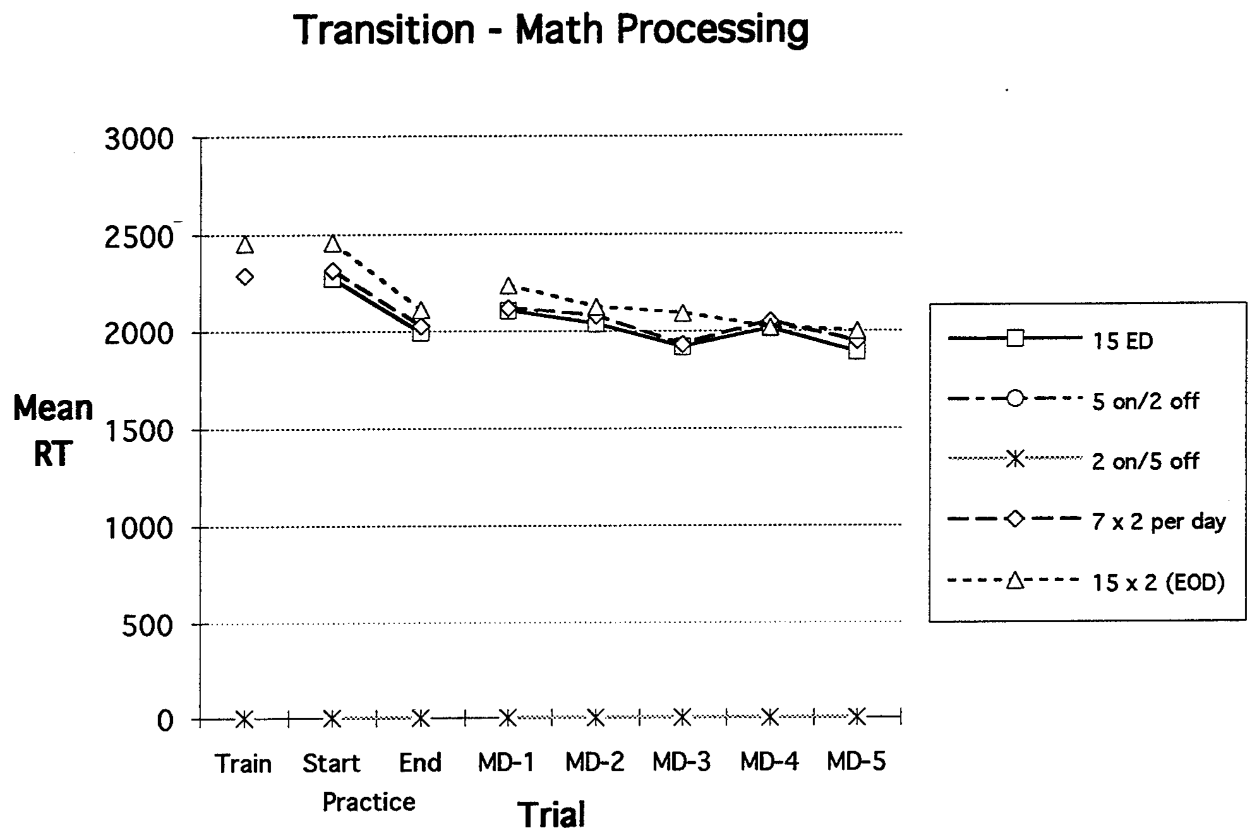
Transition - Manikin Task
Mean RT
3000
2500
2000
1500
1000
500
0
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

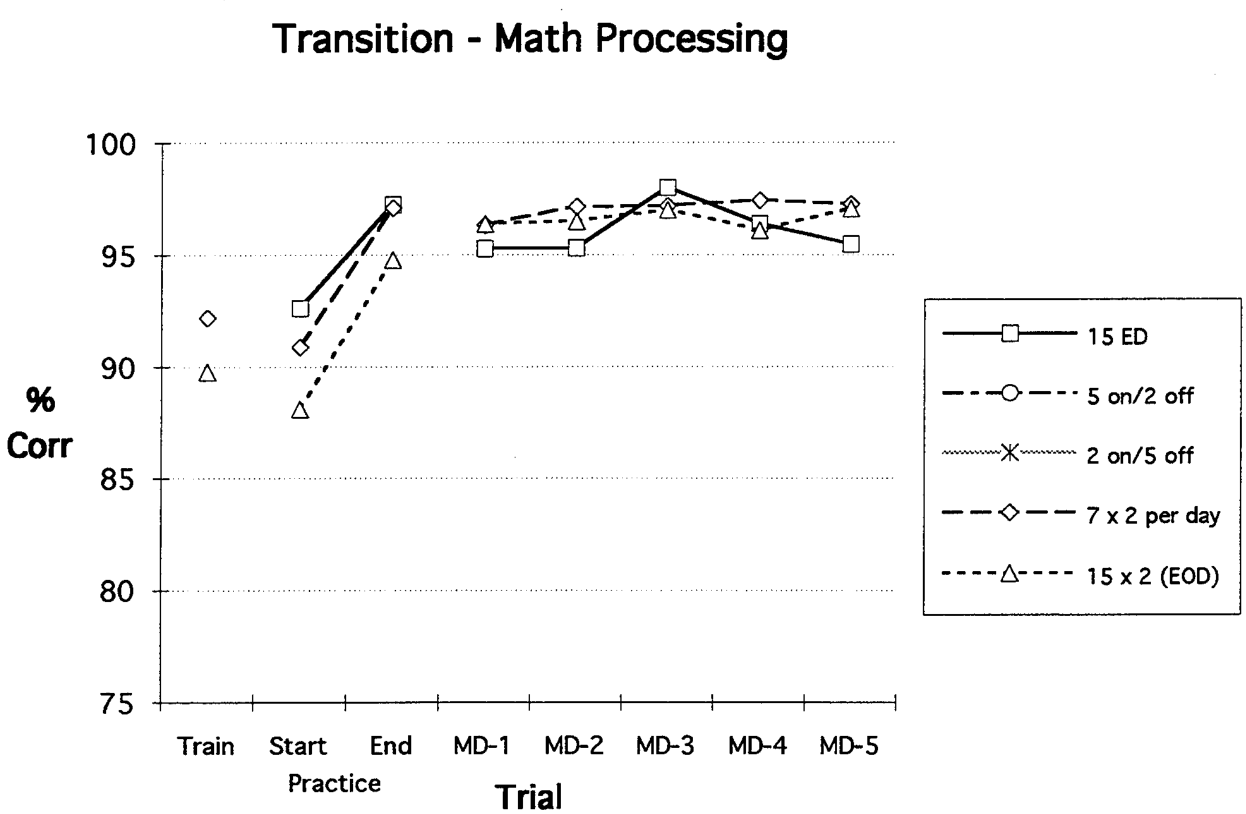


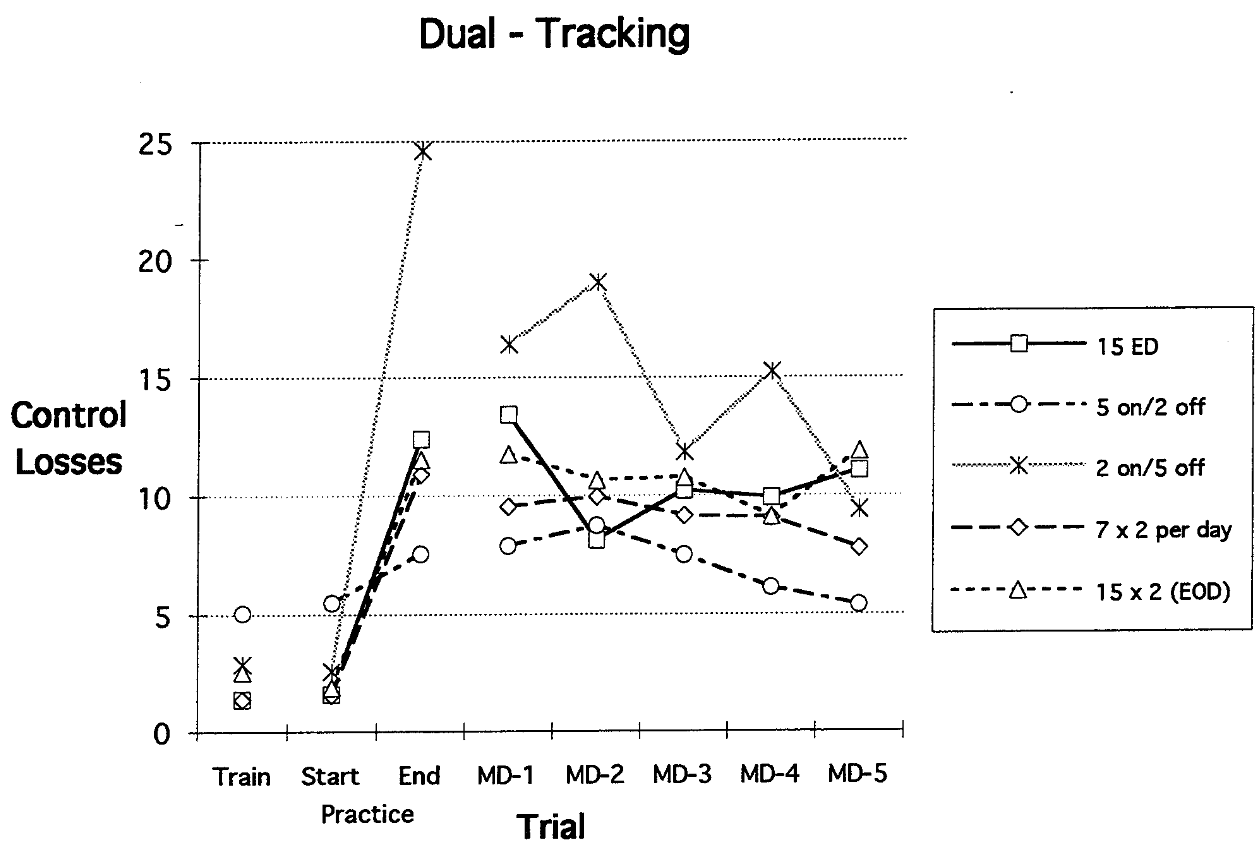
Transition - Manikin Task
% Corr
100
95
90
85
80
75
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

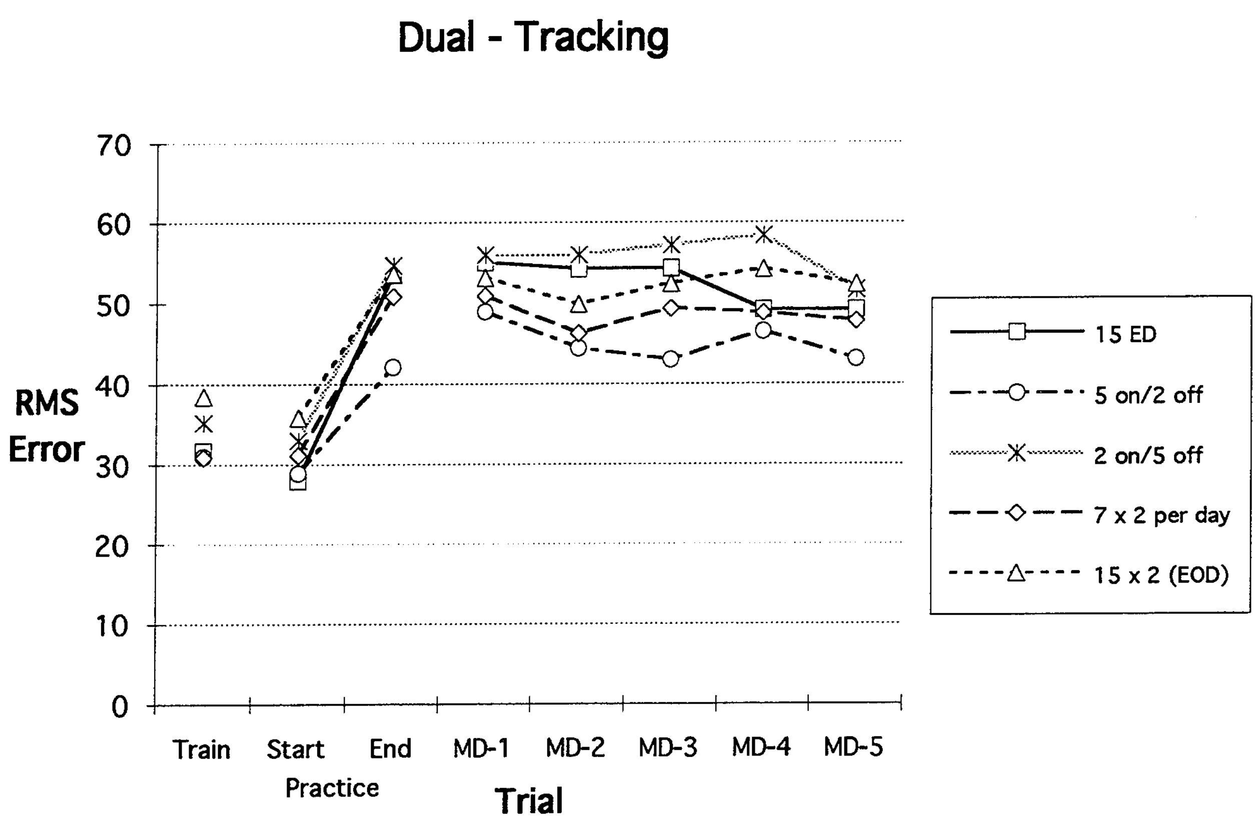

Switching - Math Processing
3000
2500
2000
1500
1000
500
0
Mean
RT
Train
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Practice
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)


Switching - Math Processing
100
95
90
85
80
75
%
Corr
Train
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Practice
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

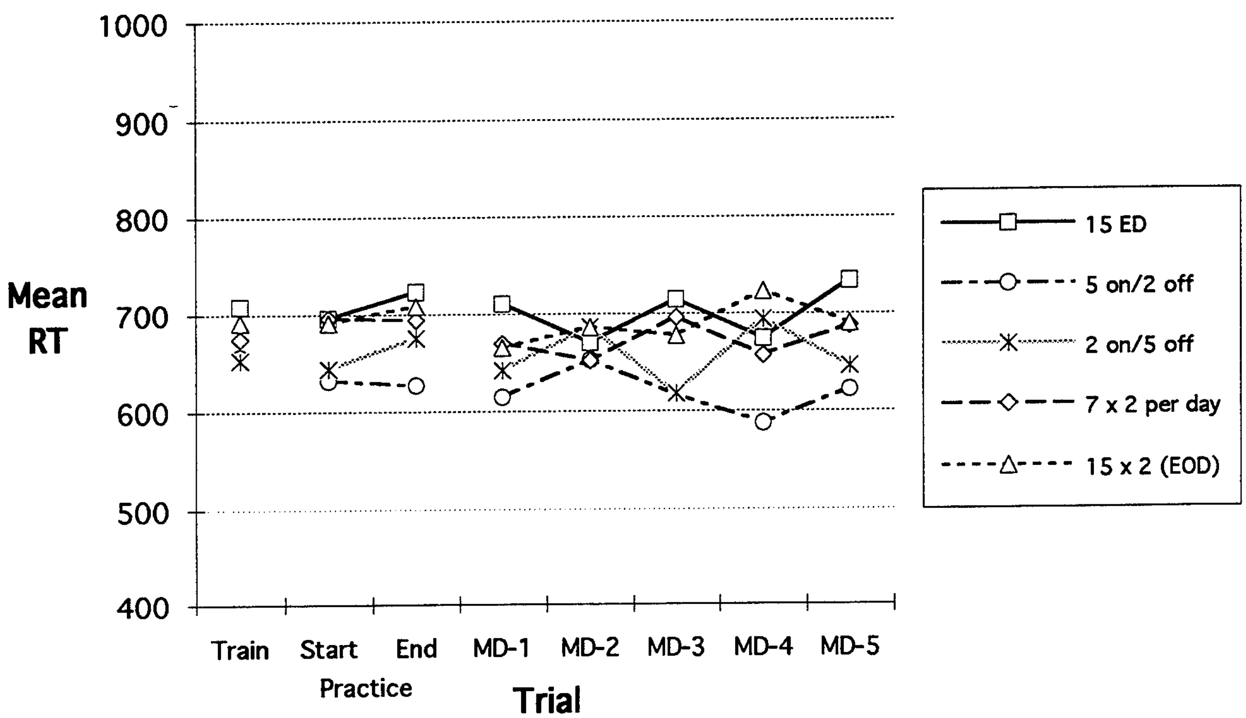

Transition - Math Processing
3000
2500
2000
1500
1000
500
0
Mean RT
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)


Transition - Math Processing
100
95
90
85
80
75
% Corr
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

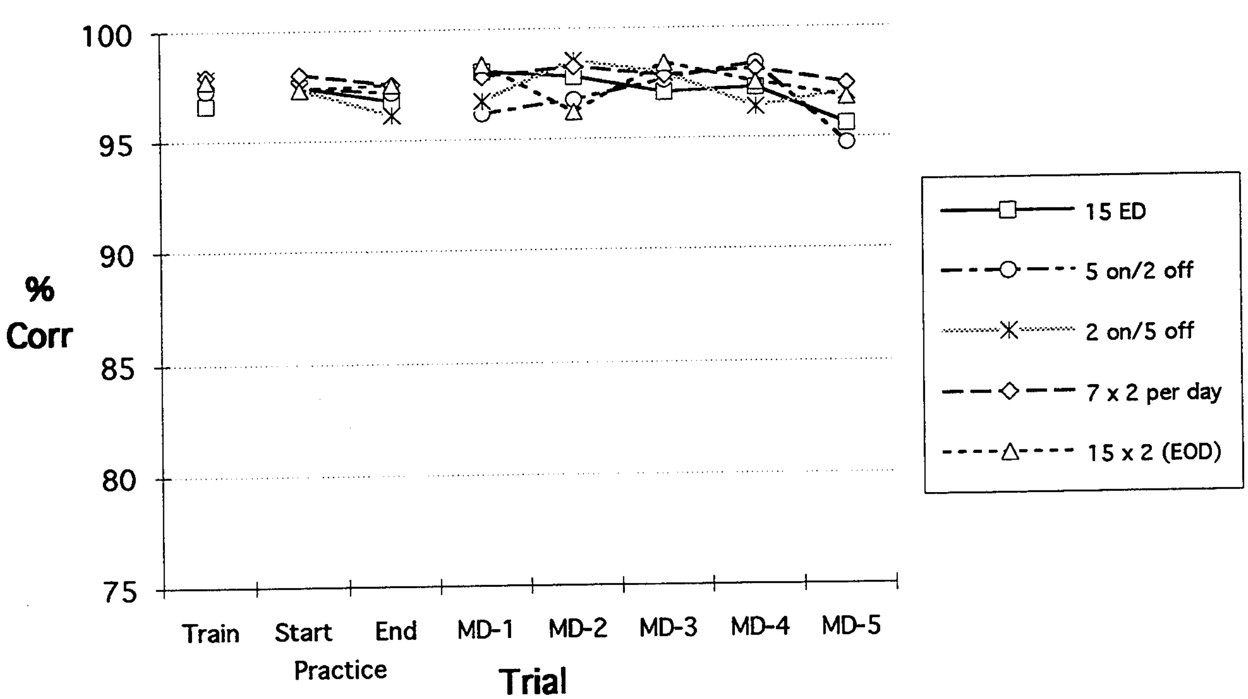

Dual - Tracking
25
20
15
10
5
0
Control Losses
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

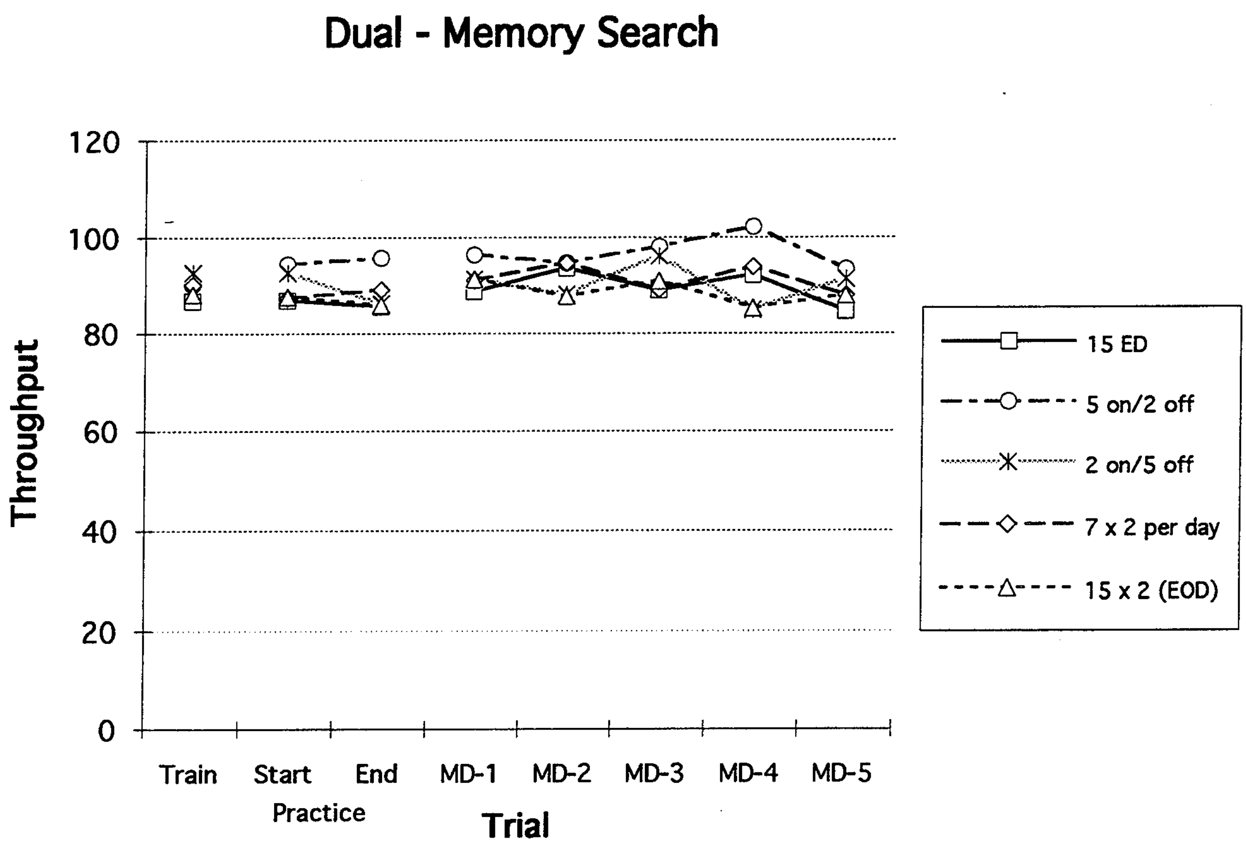

Dual - Tracking
70
60
50
40
30
20
10
0
RMS Error
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

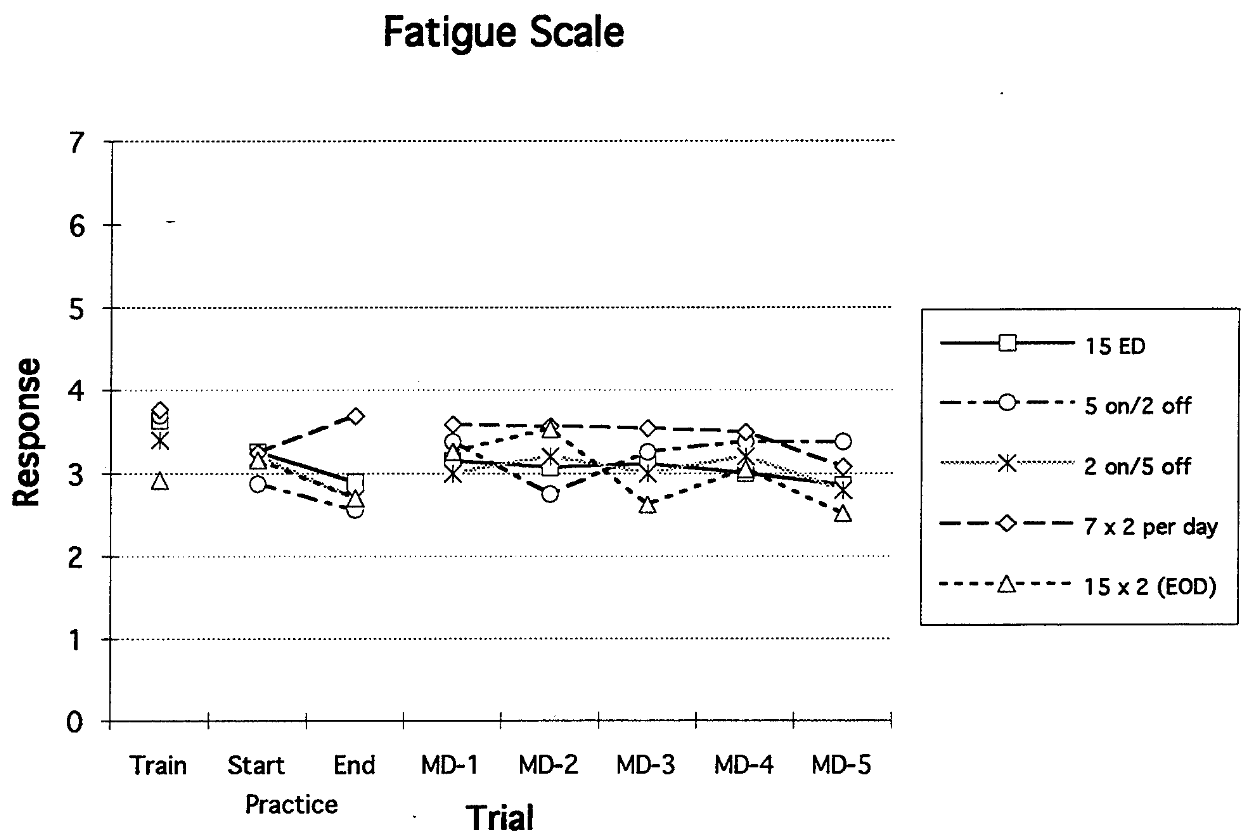
Dual - Memory Search

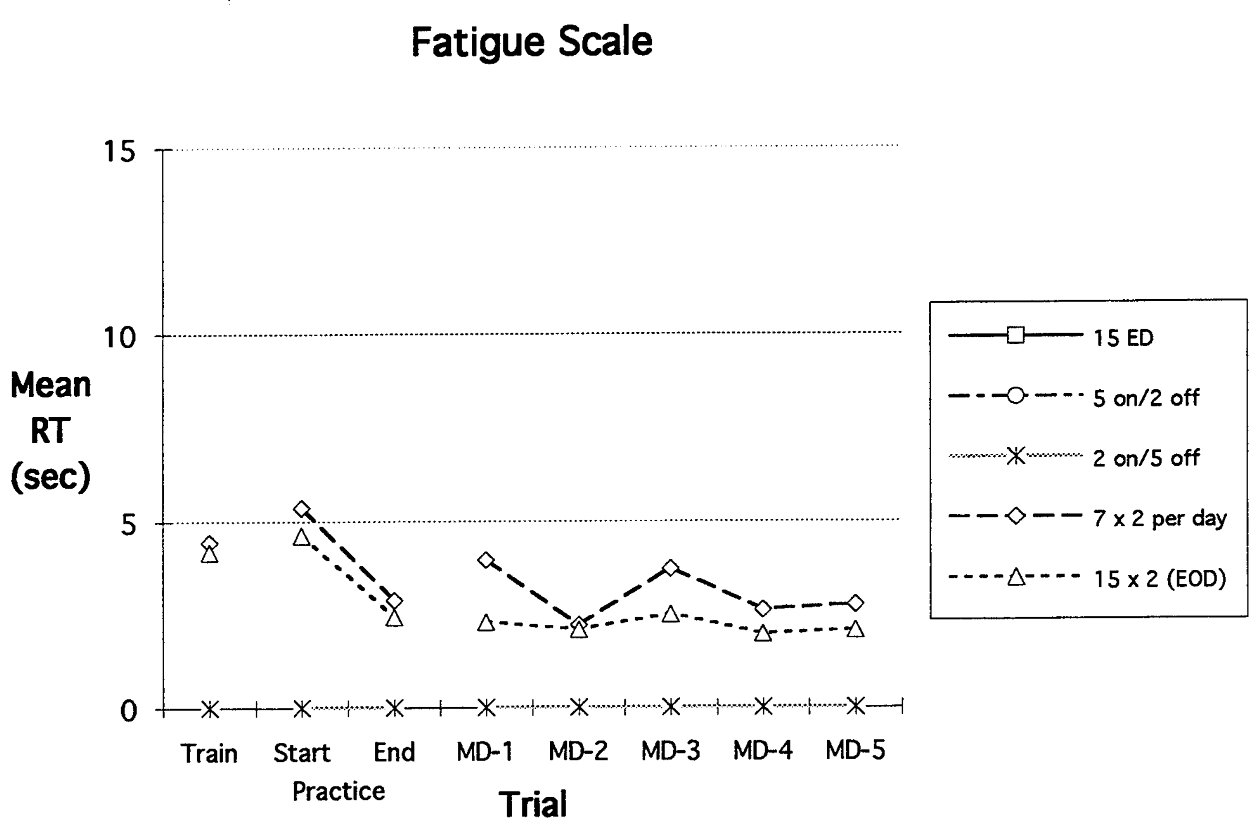


Dual - Memory Search



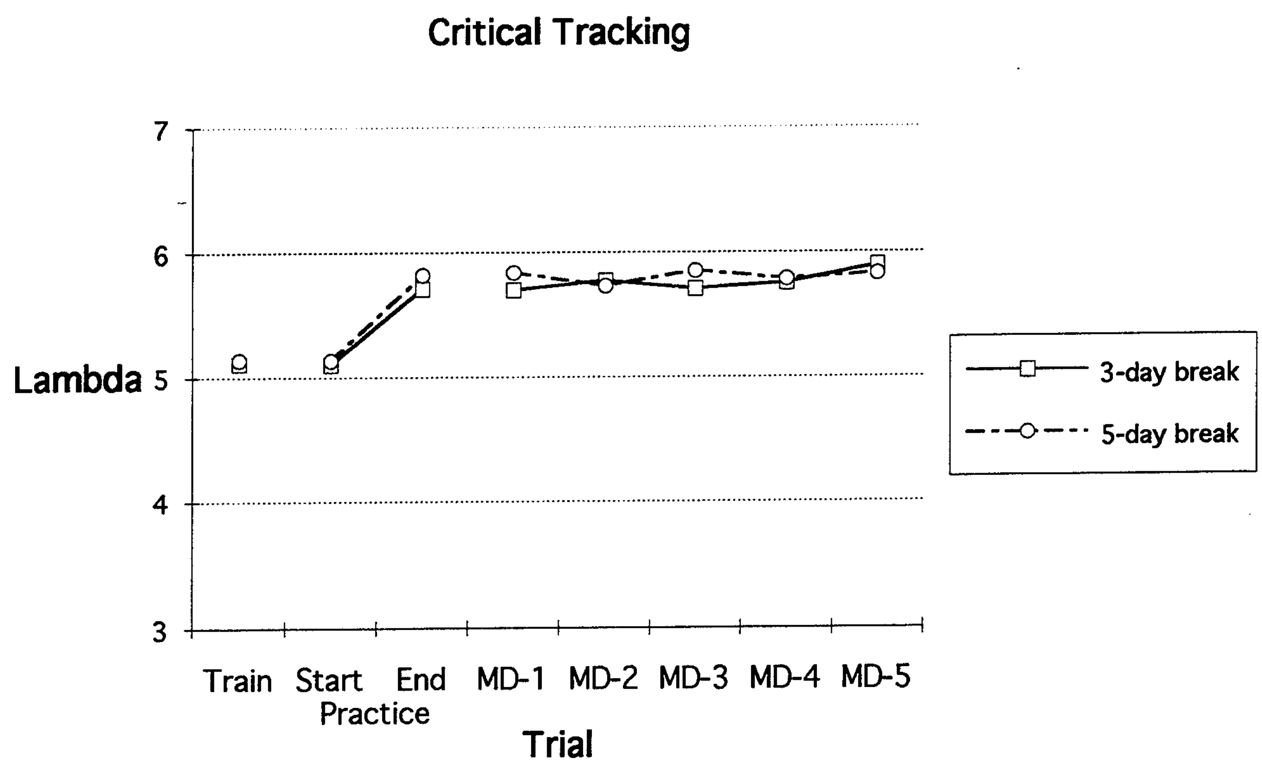

Dual - Memory Search
Throughput
120
100
80
60
40
20
0
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

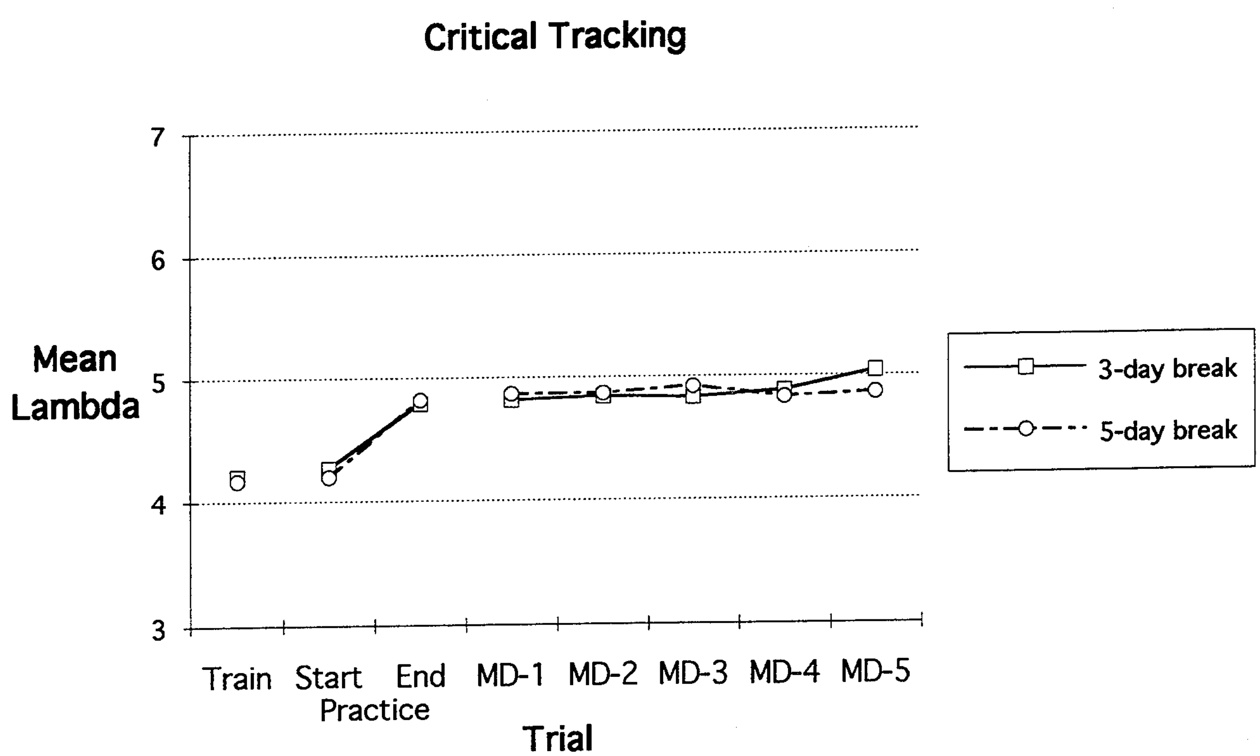

Fatigue Scale
7
6
5
4
3
2
1
0
Response
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

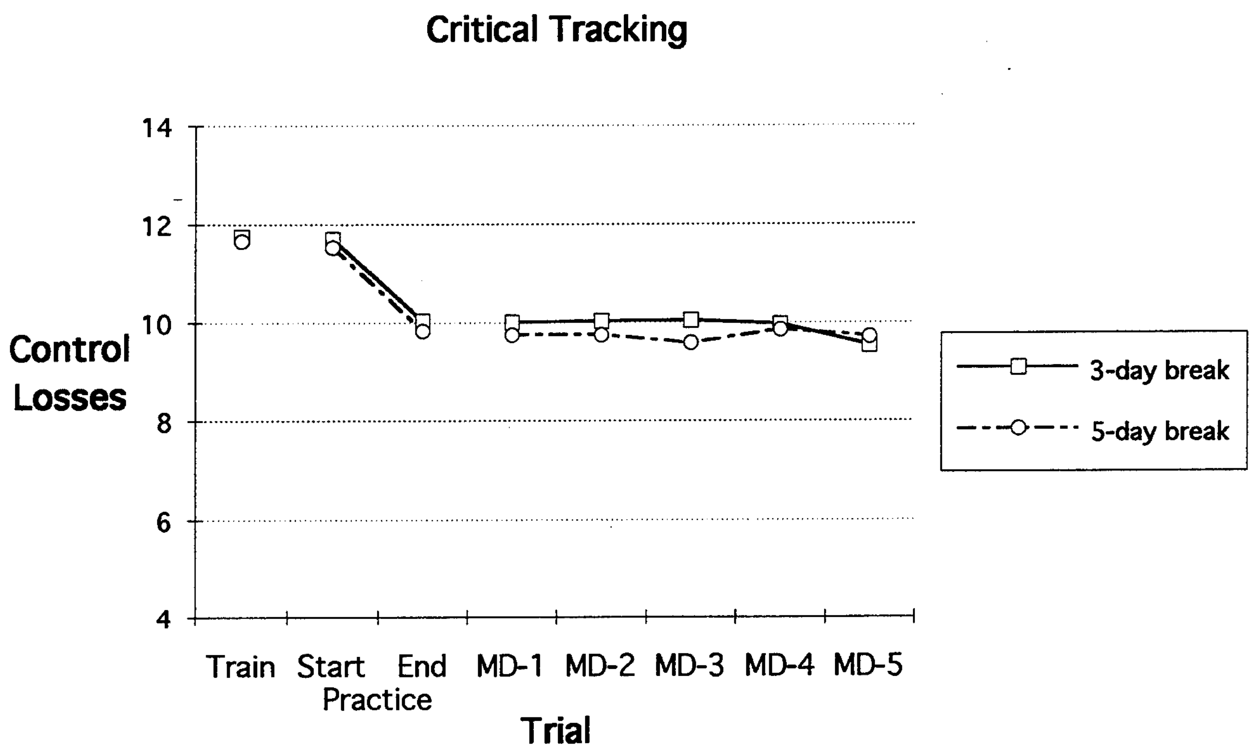

Fatigue Scale
15
10
5
0
Mean RT (sec)
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
15 ED
5 on/2 off
2 on/5 off
7 x 2 per day
15 x 2 (EOD)

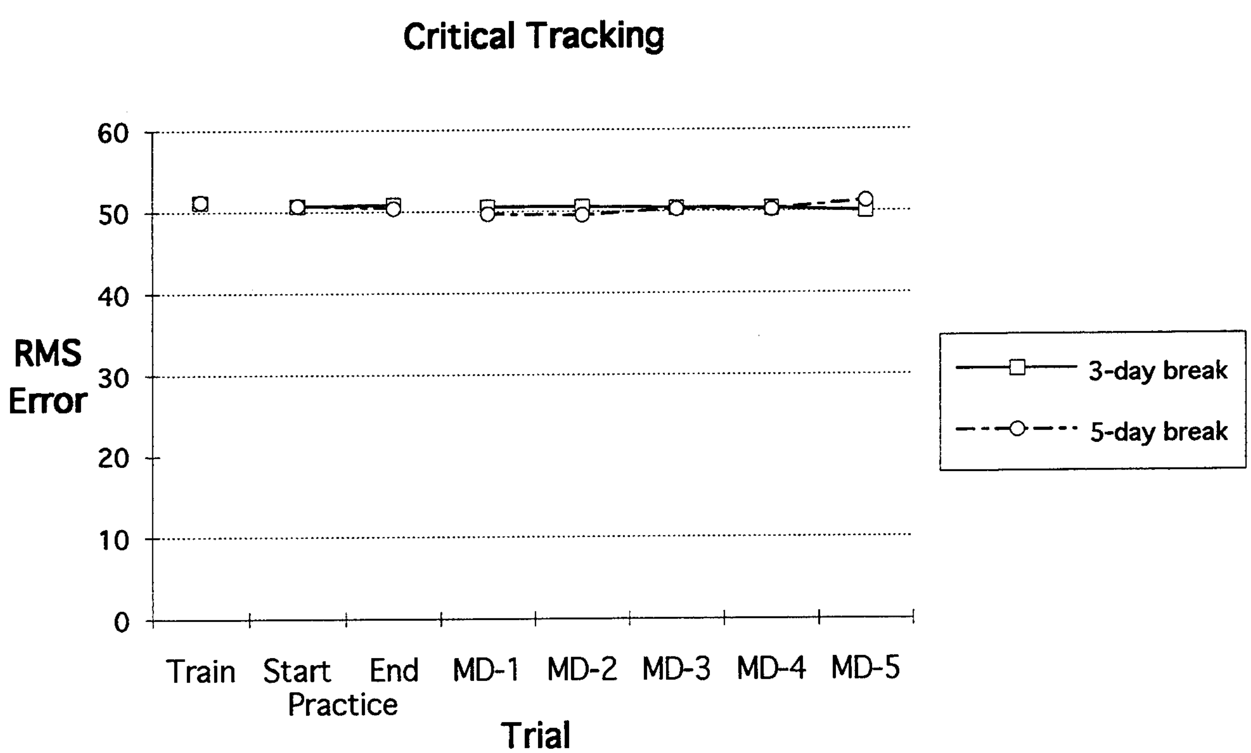
APPENDIX G

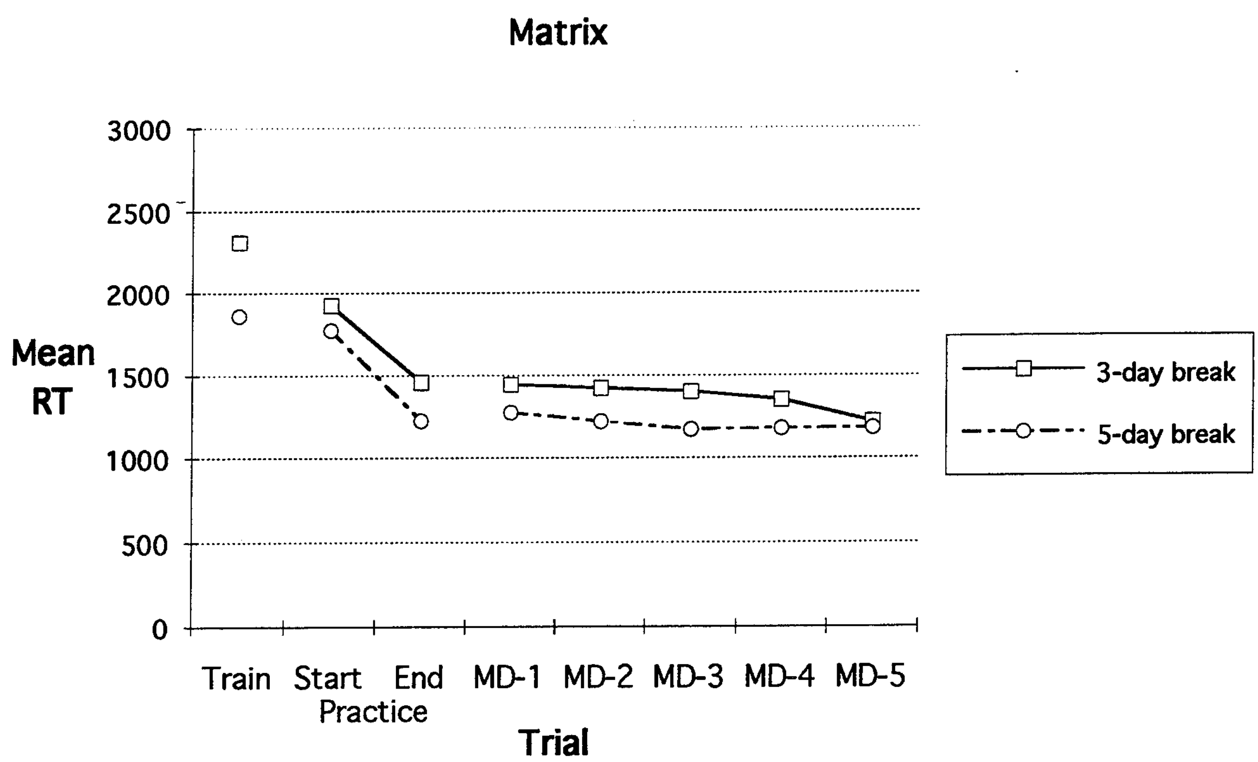
TESTING LAPSE DATA

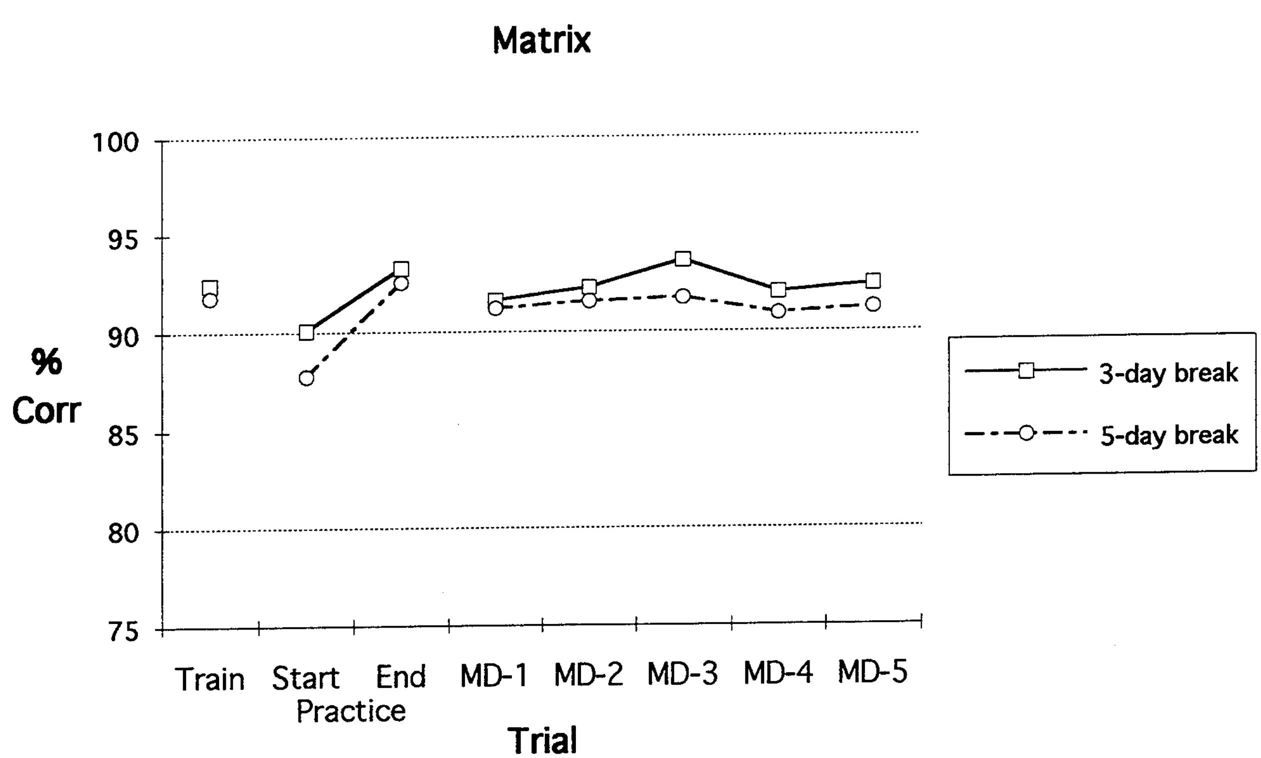

Critical Tracking
7
6
Lambda 5
4
3
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

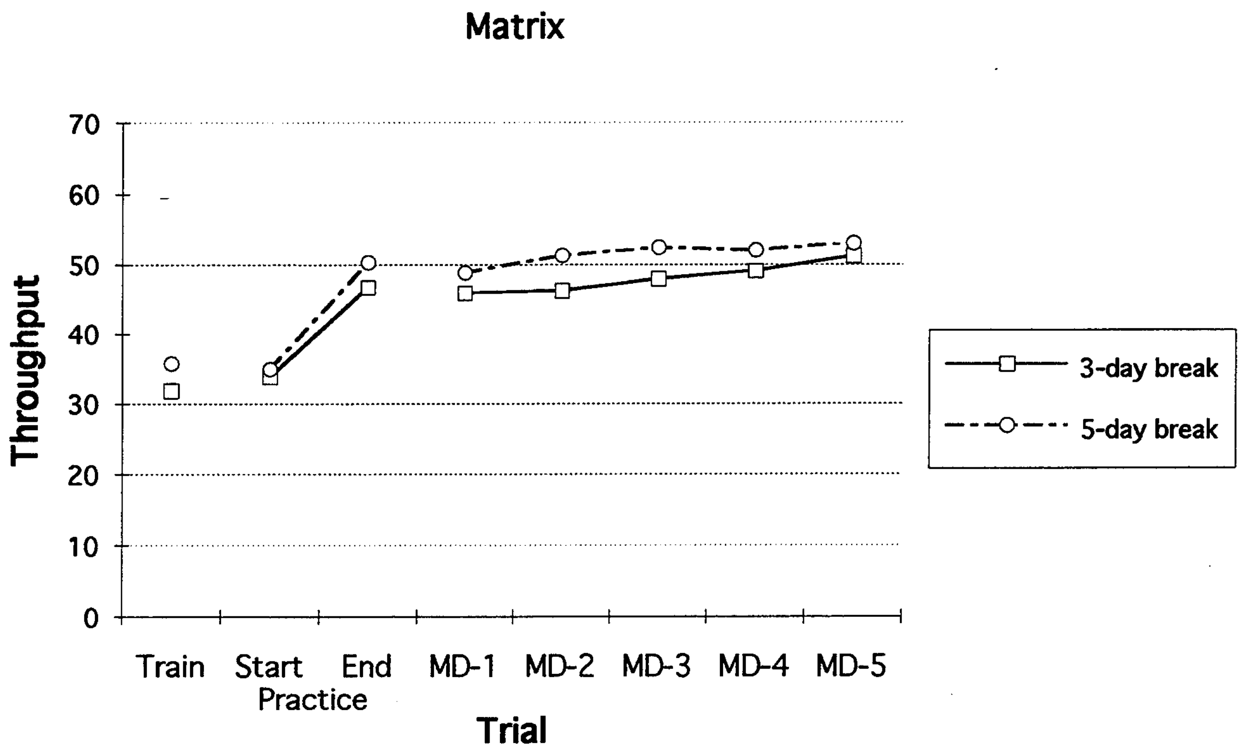

Critical Tracking
7
6
Mean Lambda 5
4
3
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

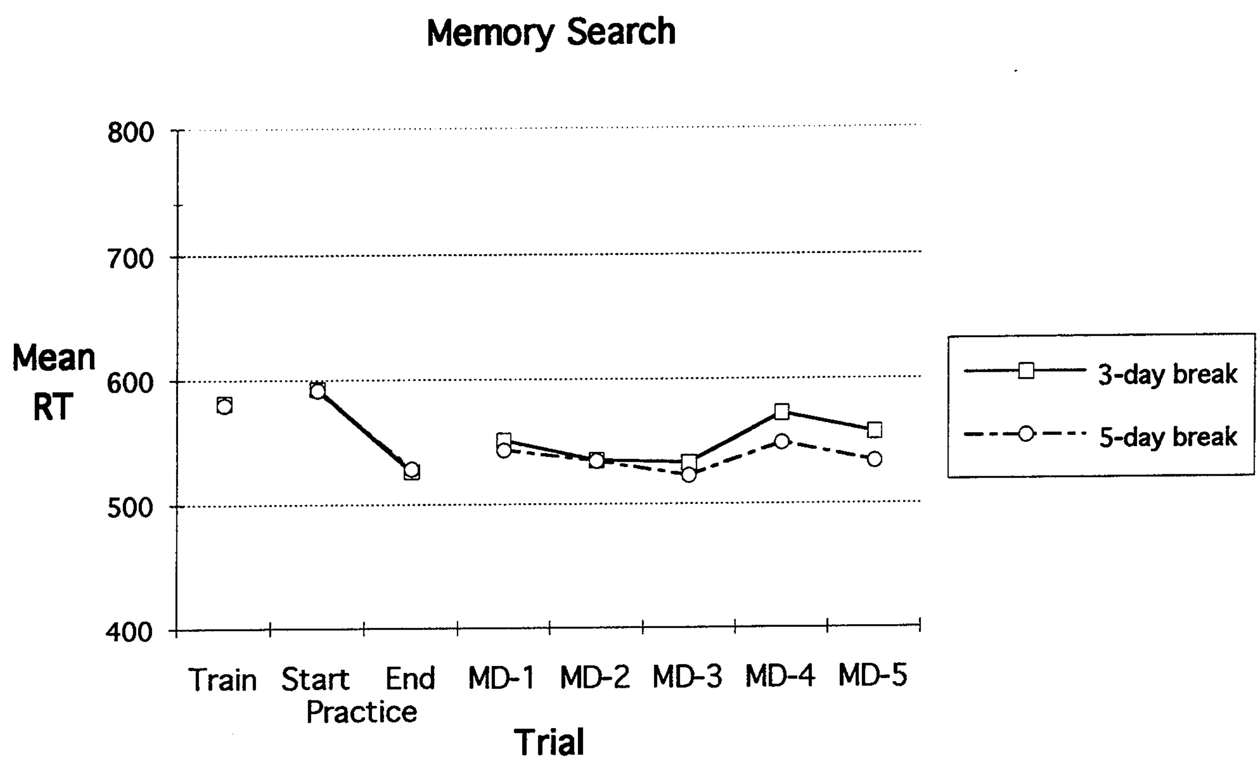

Critical Tracking
14
12
10
8
6
4
Control Losses
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

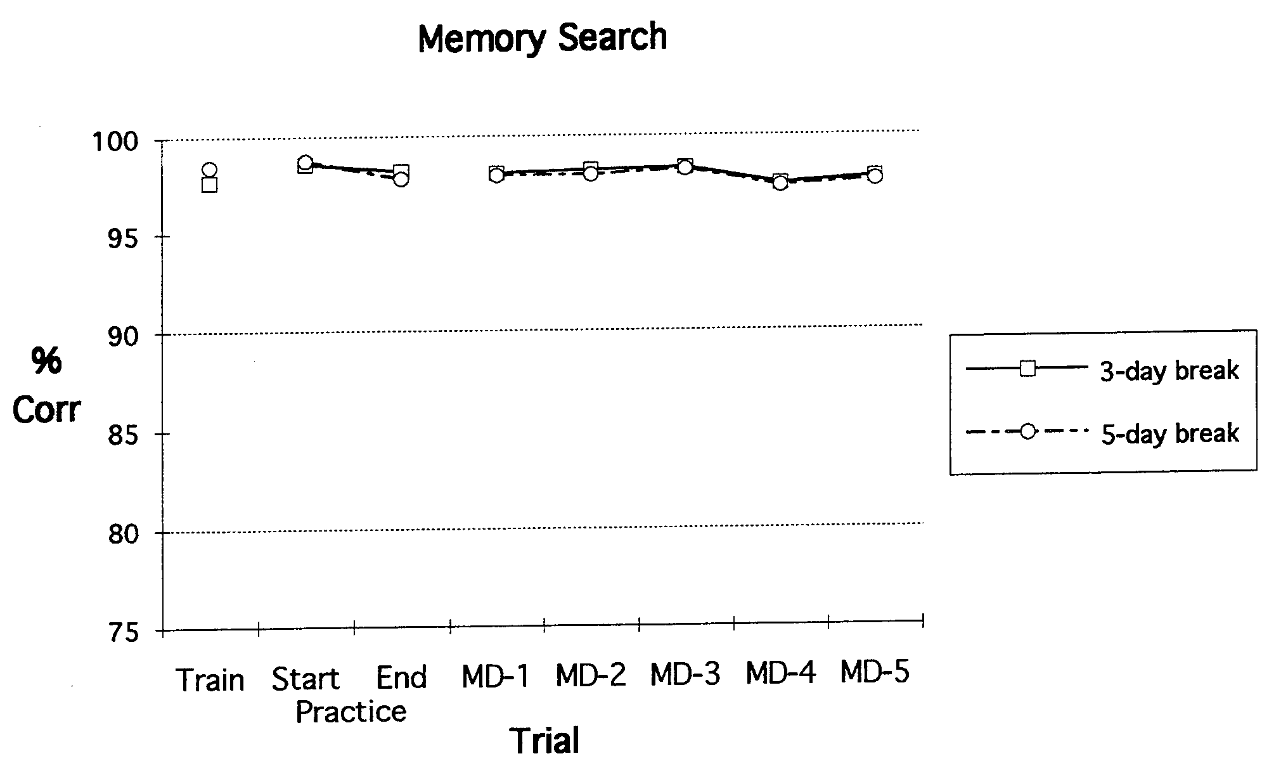

Critical Tracking
60
50
40
30
20
10
0
RMS Error
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

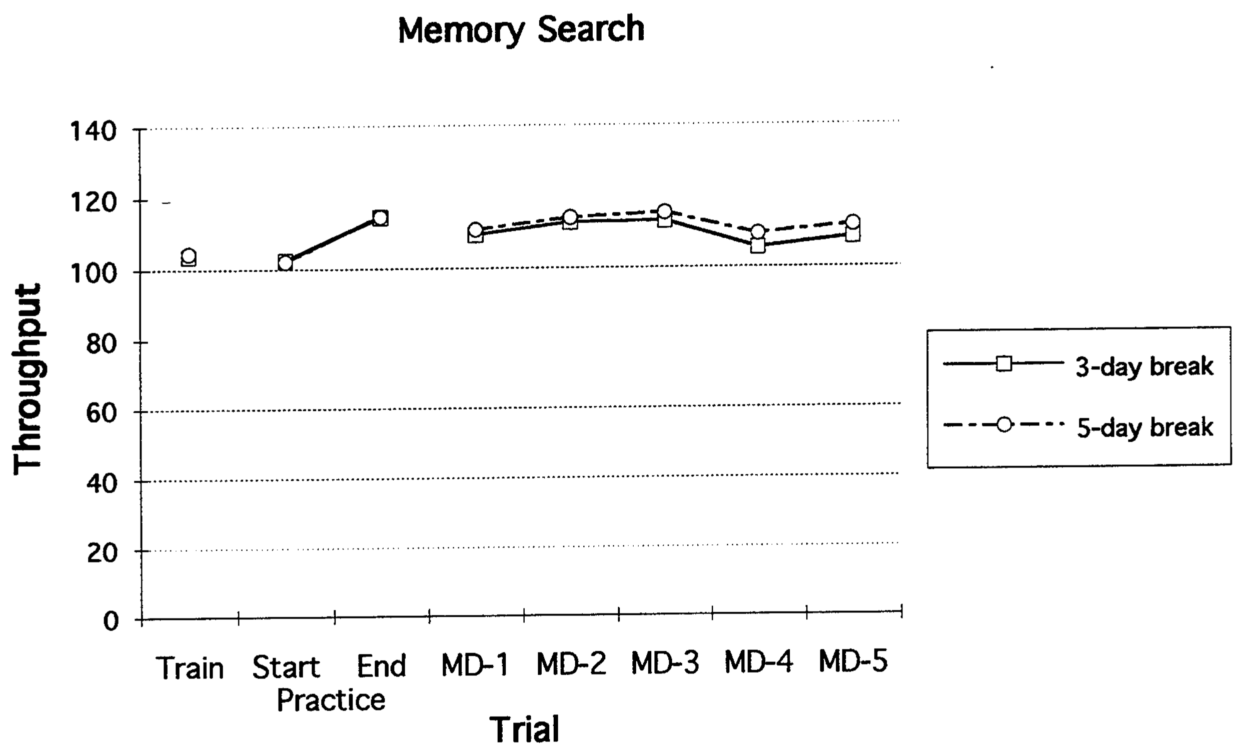

Matrix
3000
2500
2000
1500
1000
500
0
Mean RT
3-day break
5-day break
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial


Matrix
100
95
90
85
80
75
% Corr
3-day break
5-day break
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

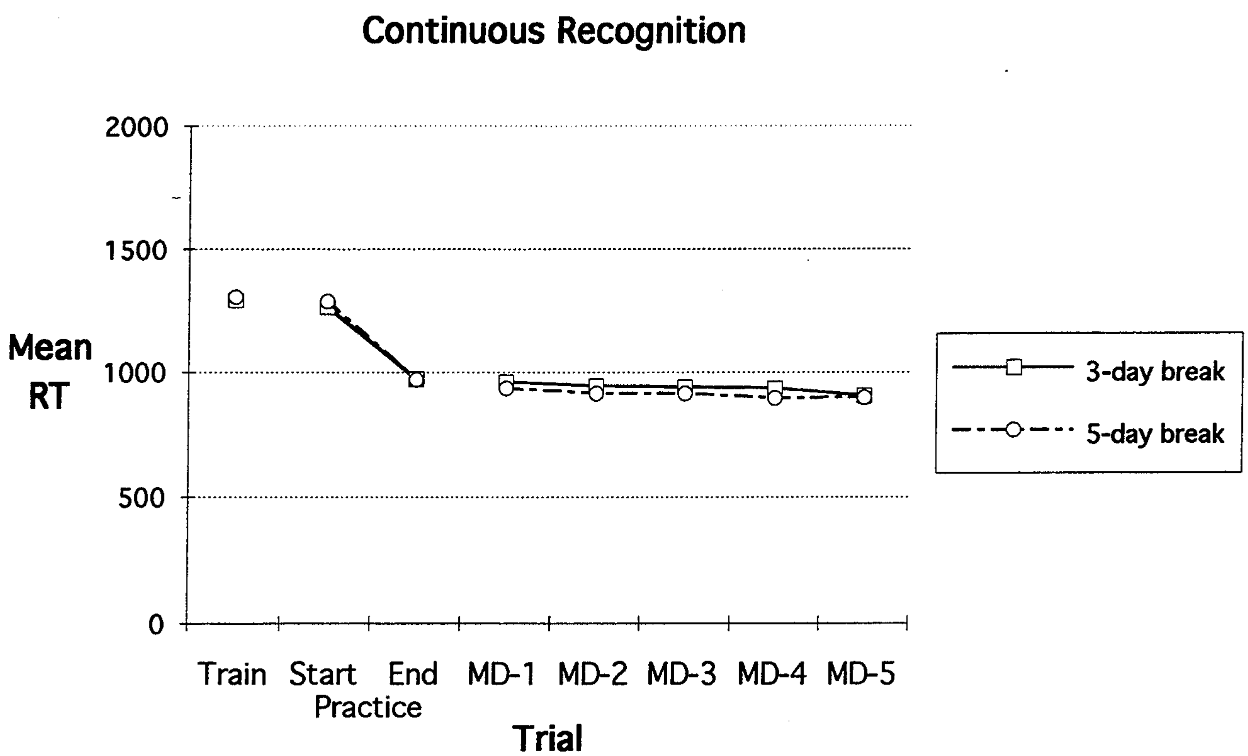

Matrix
70
60
50
40
30
20
10
0
Throughput
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

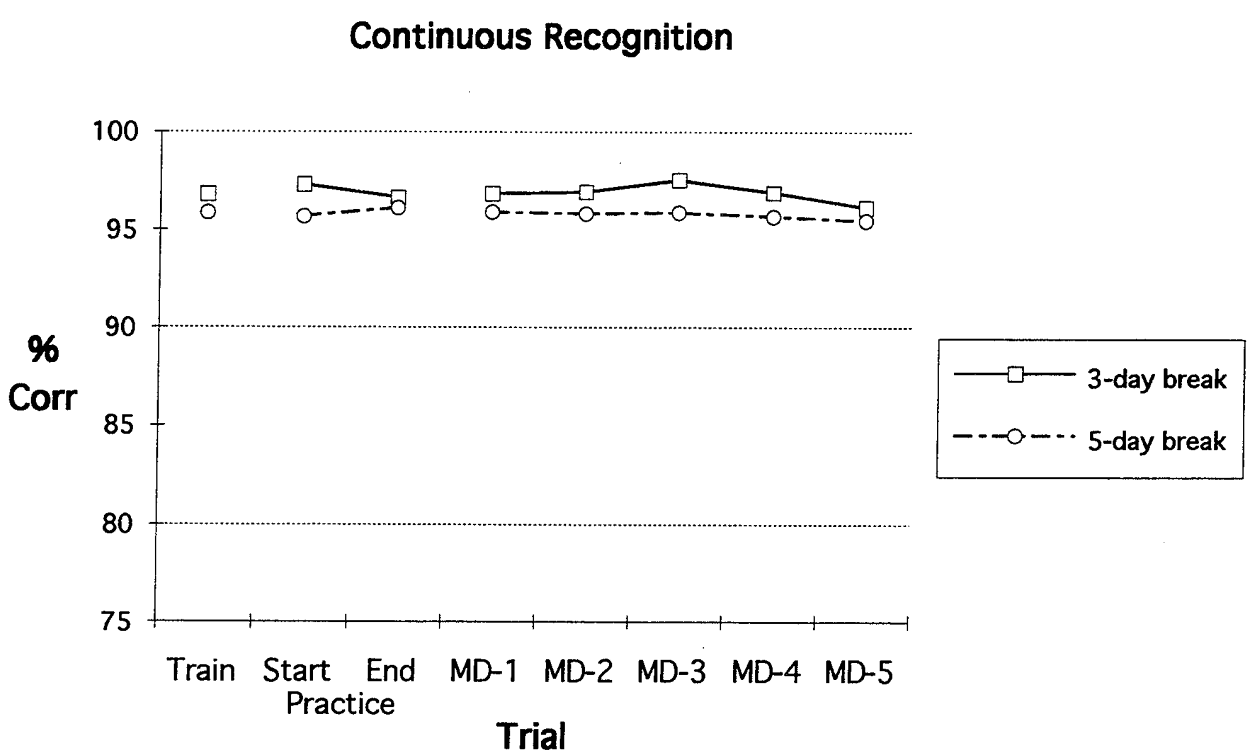

Memory Search
Mean RT
800
700
600
500
400
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

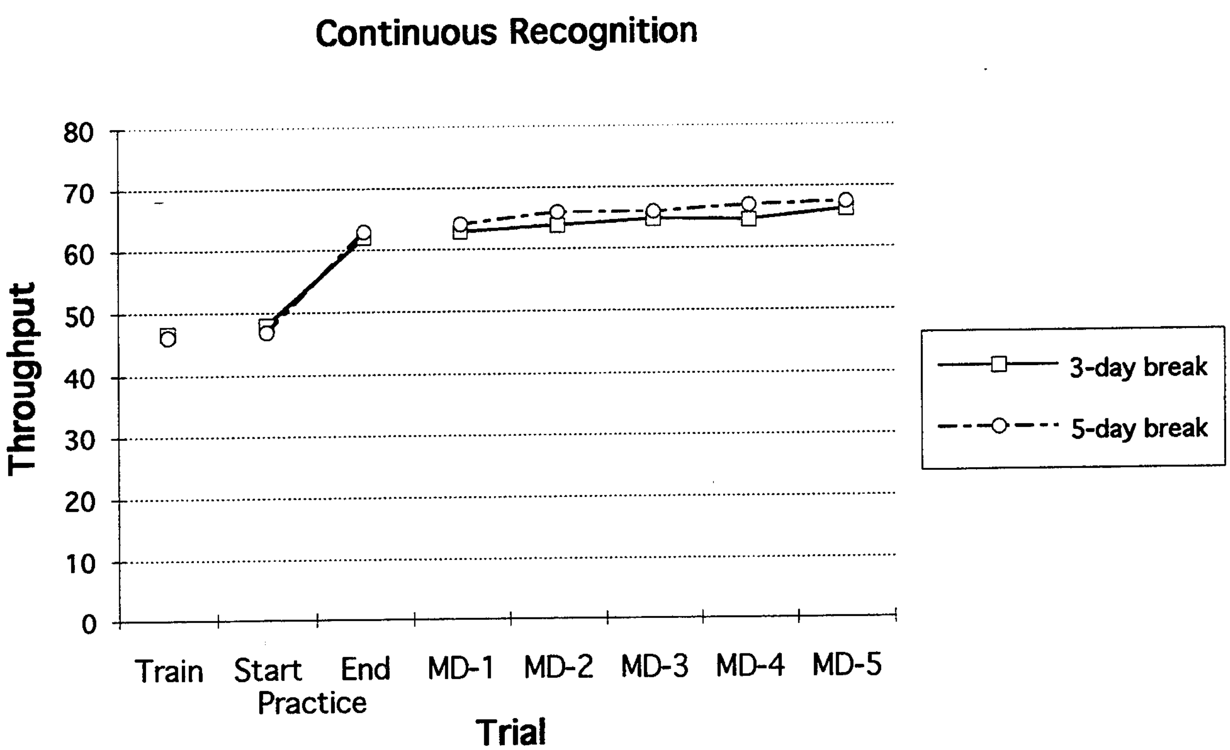

Memory Search
% Corr
100
95
90
85
80
75
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

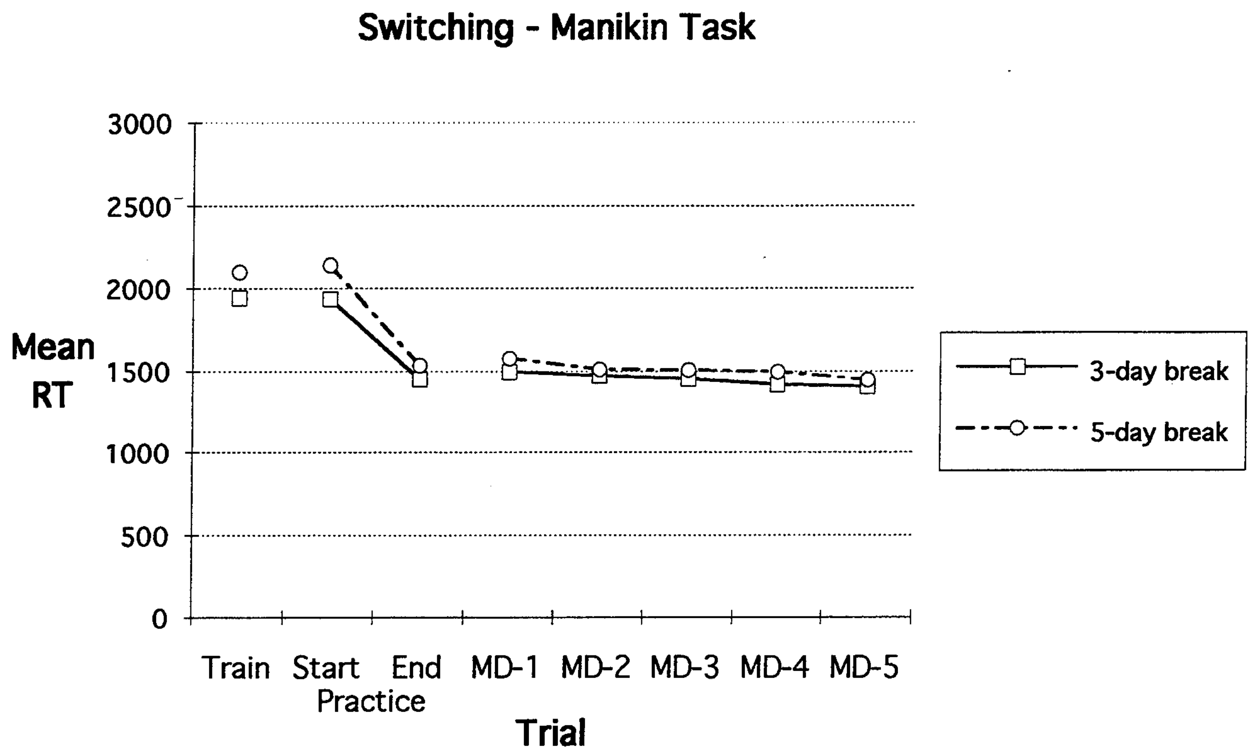


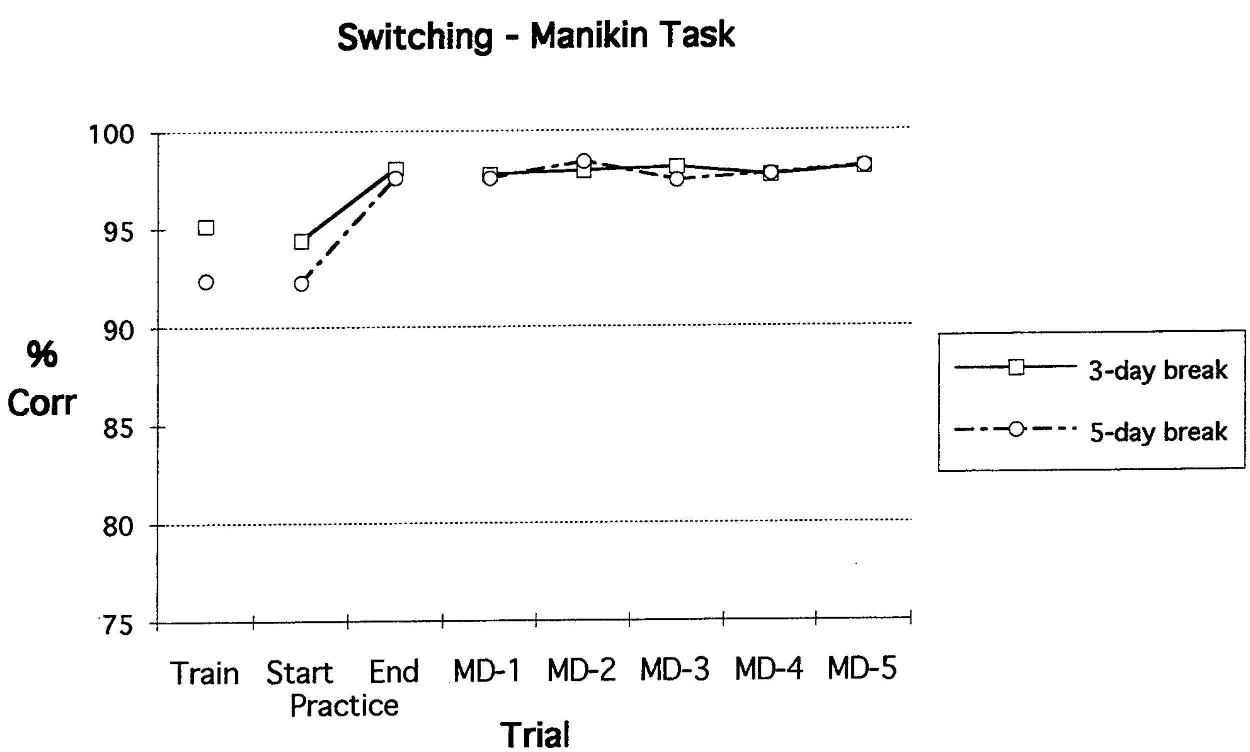
Memory Search
140
120
100
80
60
40
20
0
Throughput
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

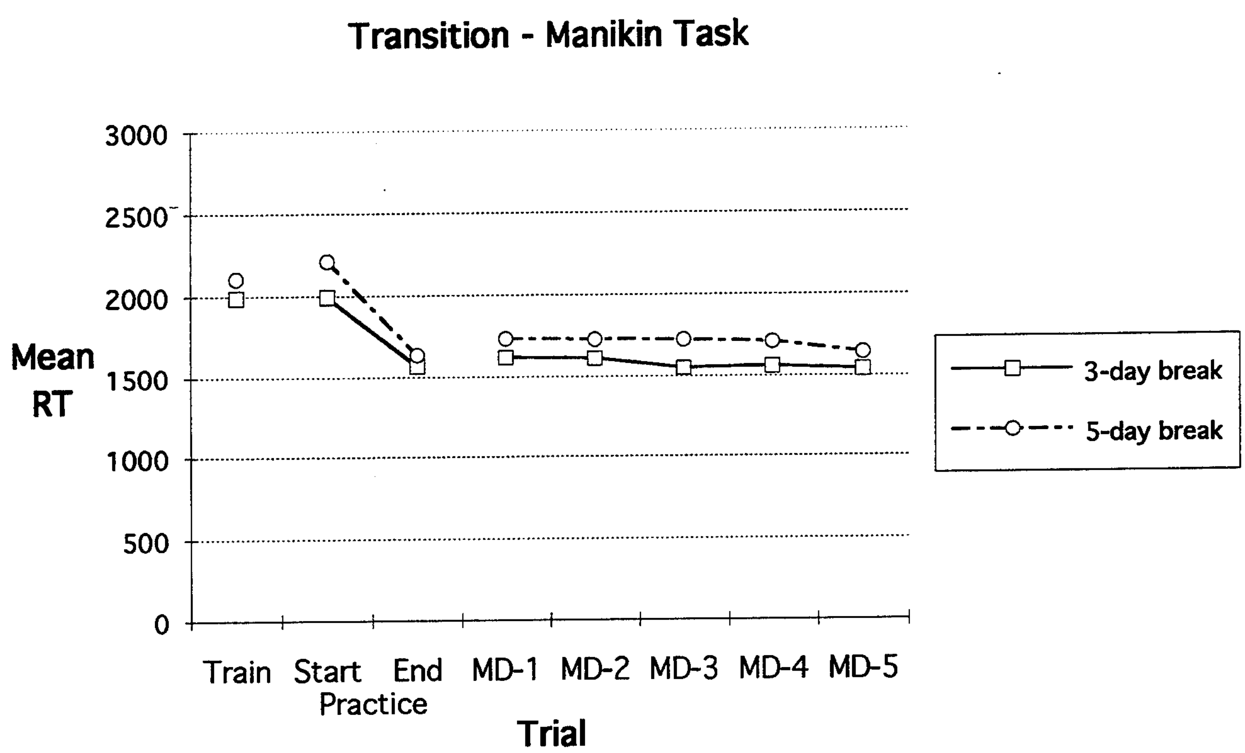

Continuous Recognition
2000
1500
1000
500
0
Mean
RT
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

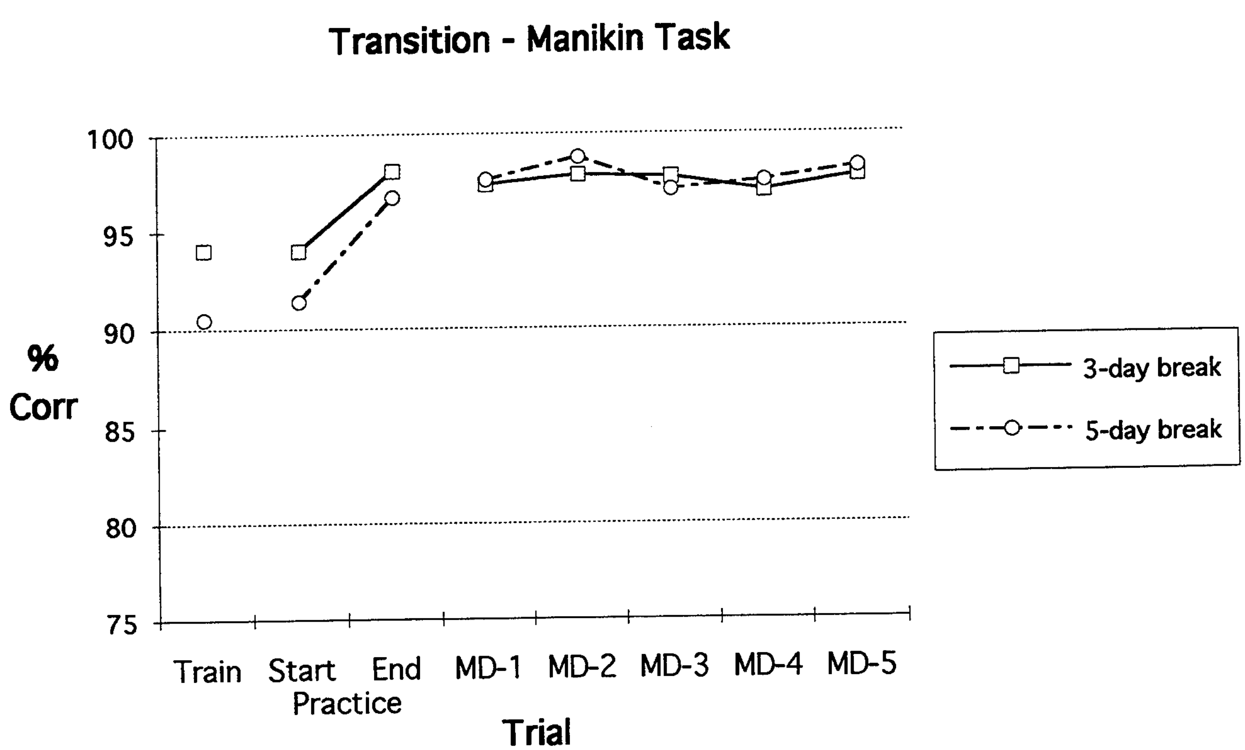

Continuous Recognition
100
95
90
85
80
75
%
Corr
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break


Continuous Recognition
80
70
60
50
40
30
20
10
0
Throughput
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

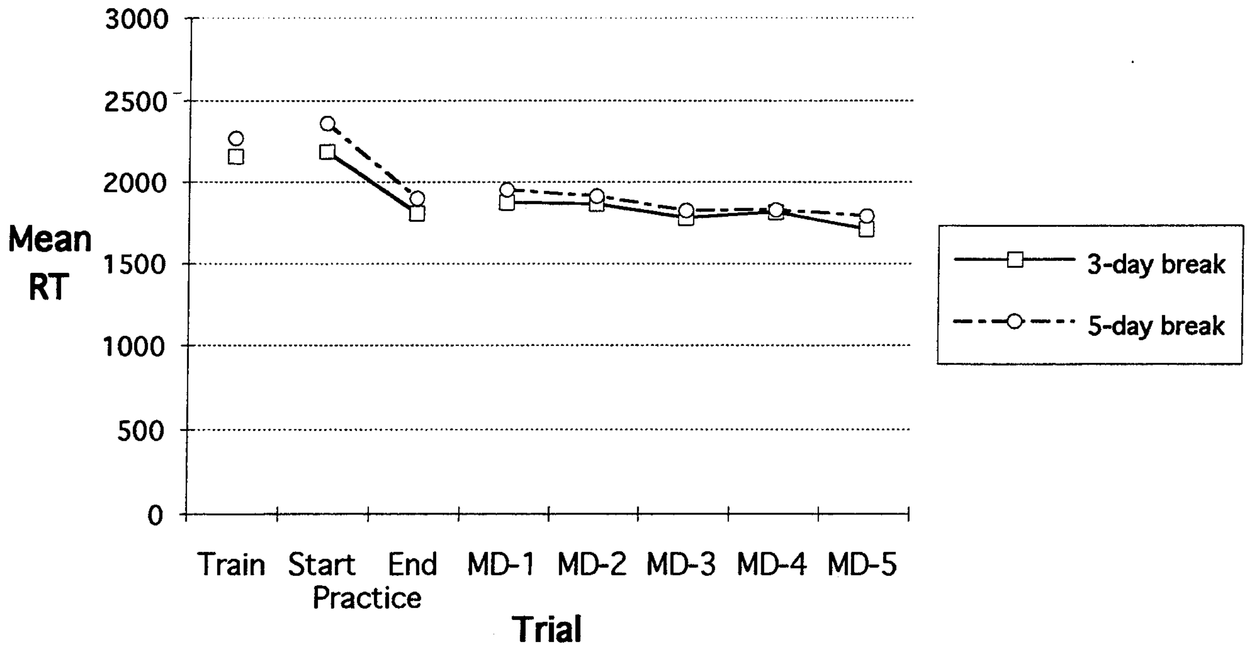

Switching - Manikin Task
Mean RT
3000
2500
2000
1500
1000
500
0
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break


Switching - Manikin Task
% Corr
100
95
90
85
80
75
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

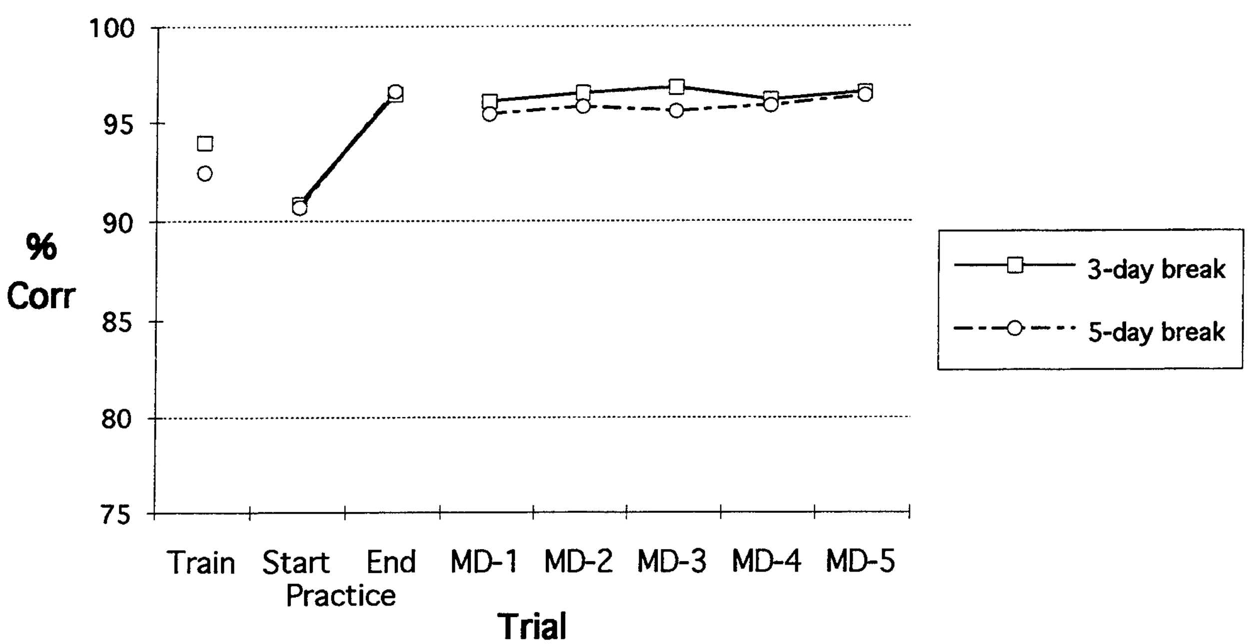

Transition - Manikin Task
3000
2500
2000
1500
1000
500
0
Mean RT
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break


Transition - Manikin Task
100
95
90
85
80
75
% Corr
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

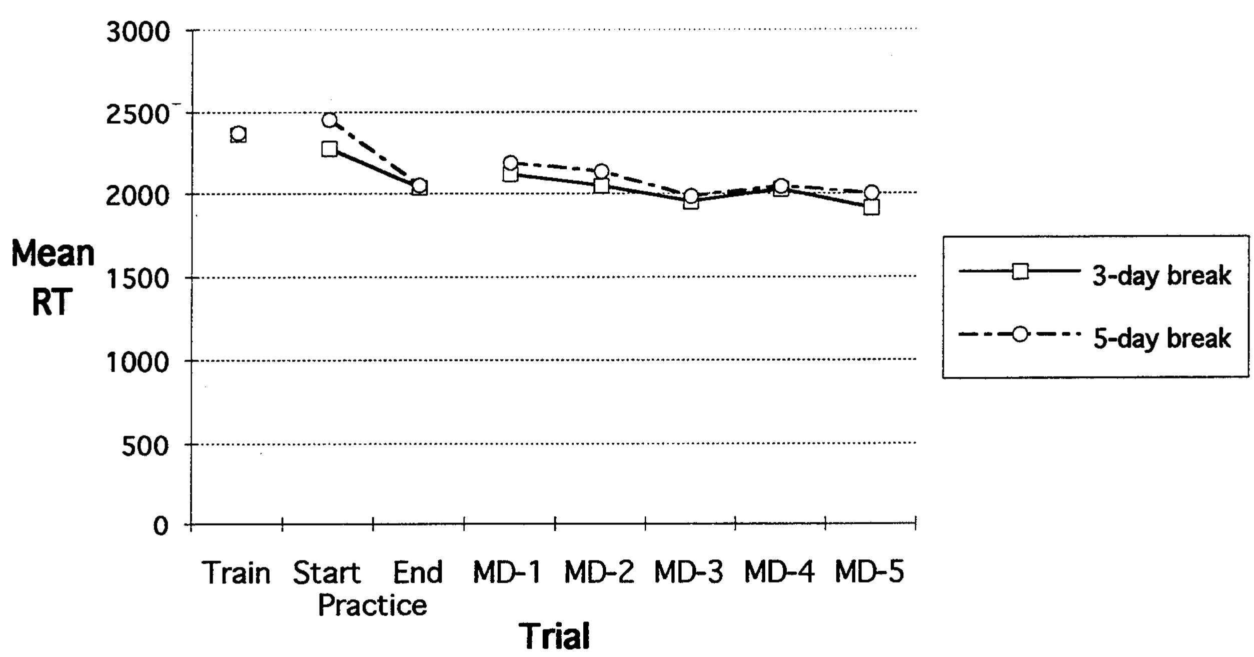
Switching - Math Processing



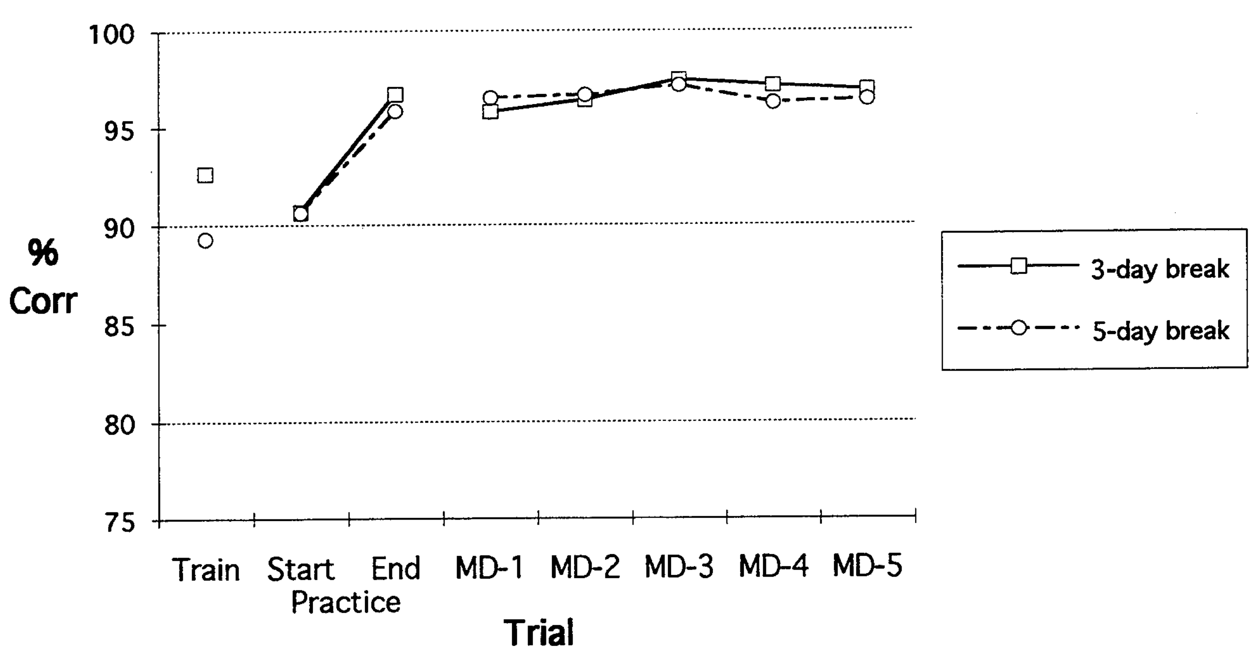
Switching - Math Processing

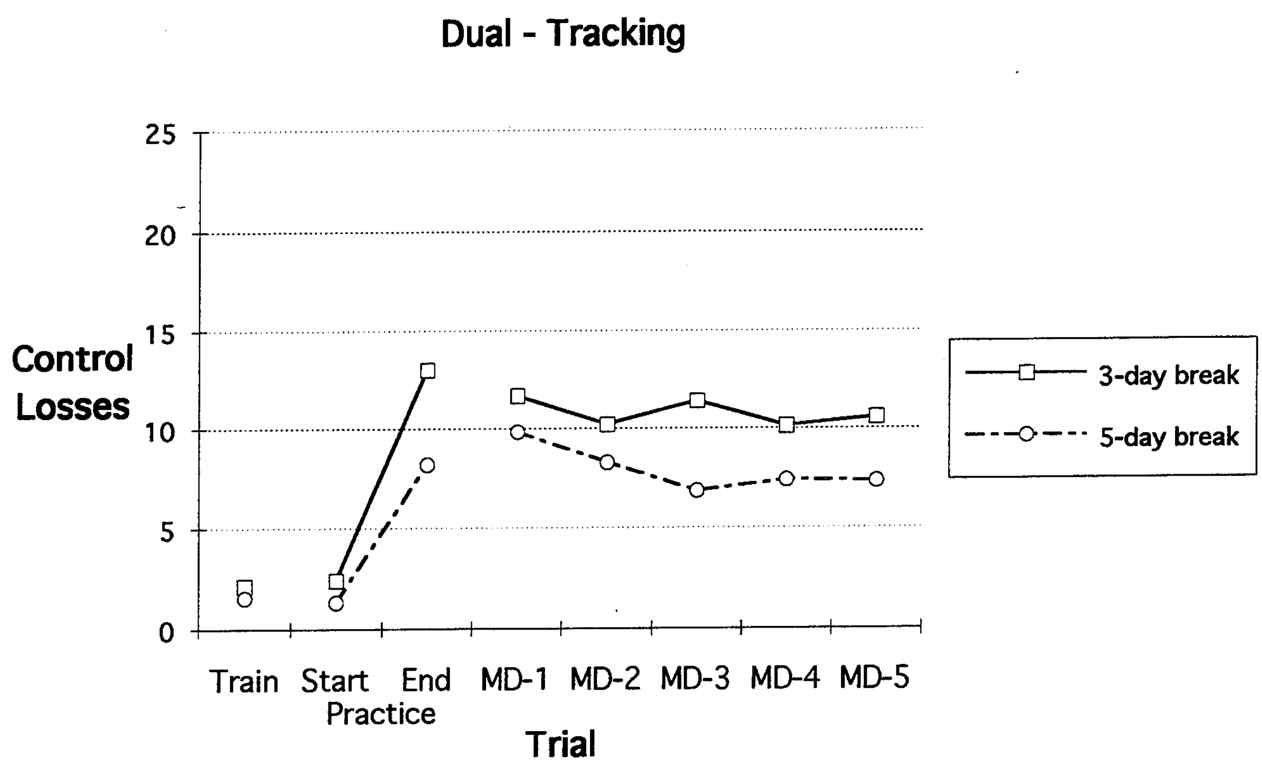


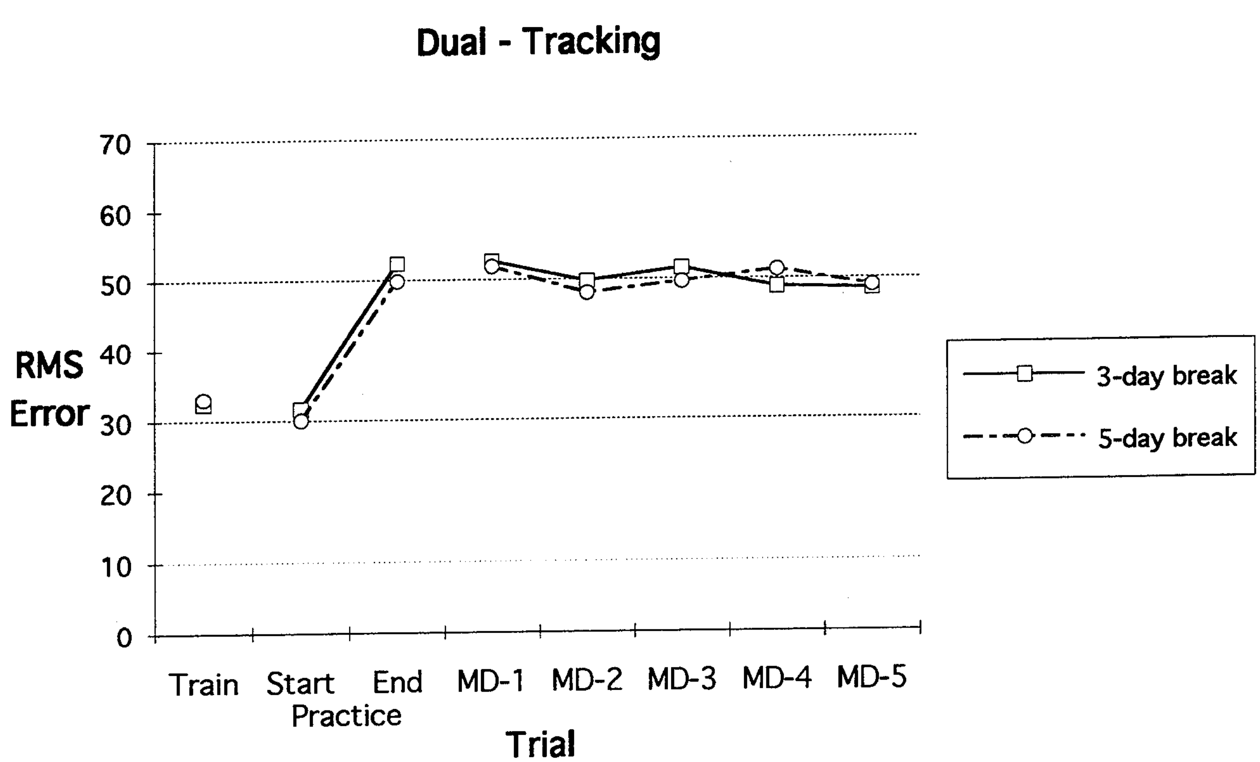
Transition - Math Processing



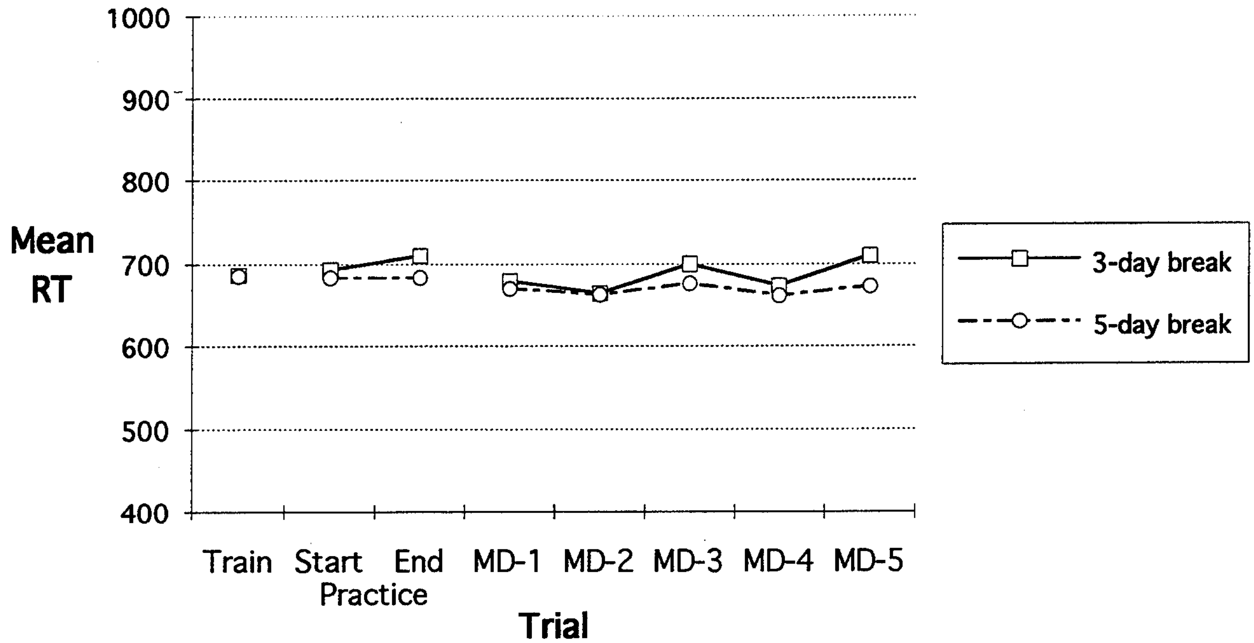
Transition - Math Processing



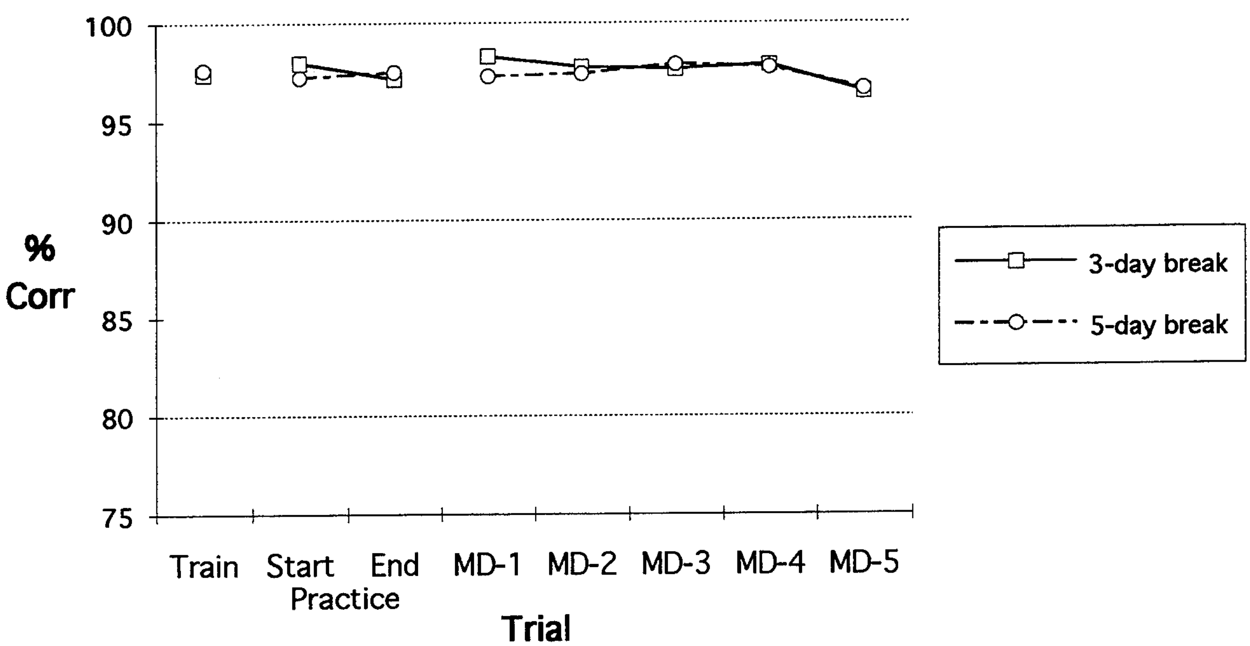

Dual - Tracking
25
20
15
10
5
0
Control Losses
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

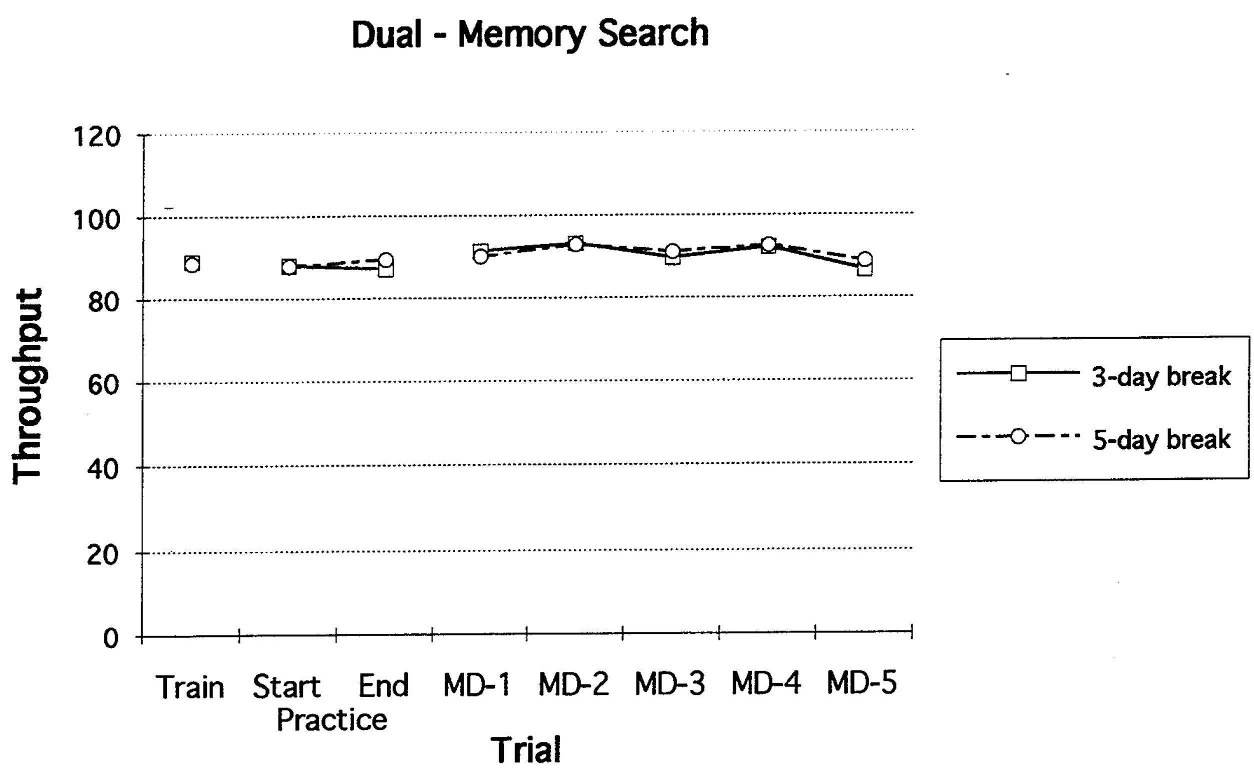

Dual - Tracking
70
60
50
40
30
20
10
0
RMS Error
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

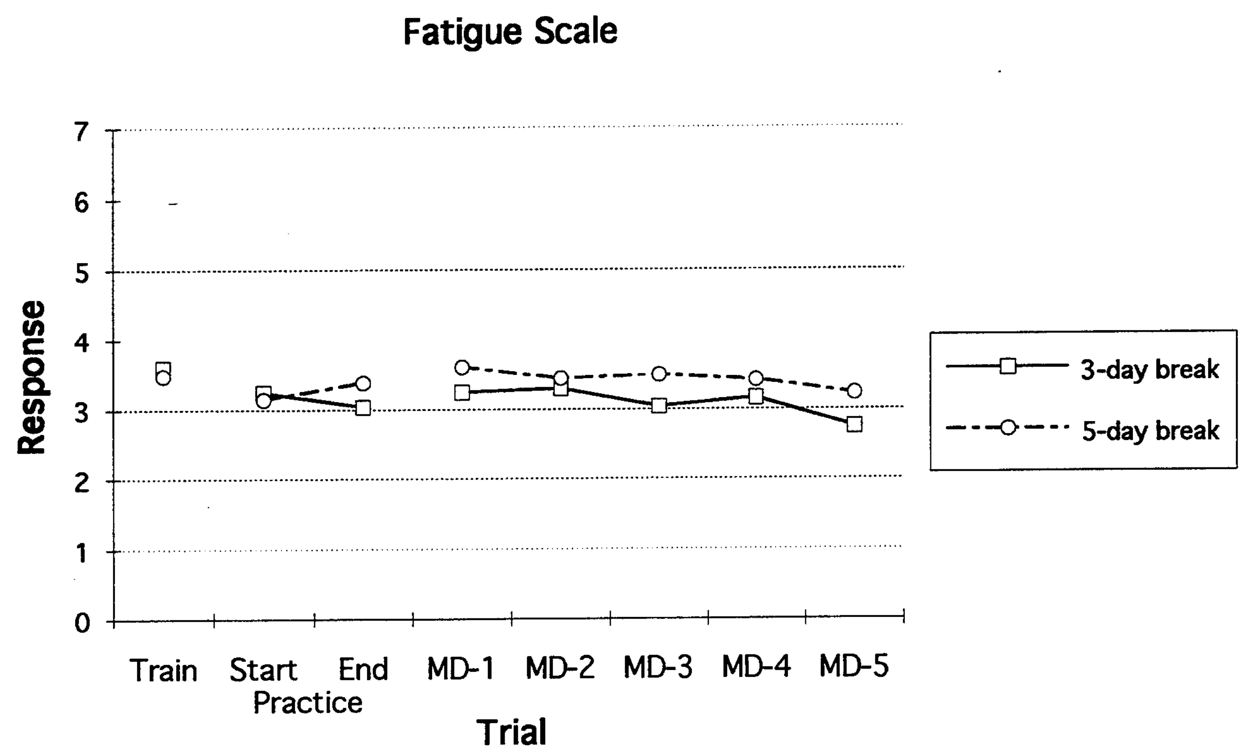
Dual - Memory Search

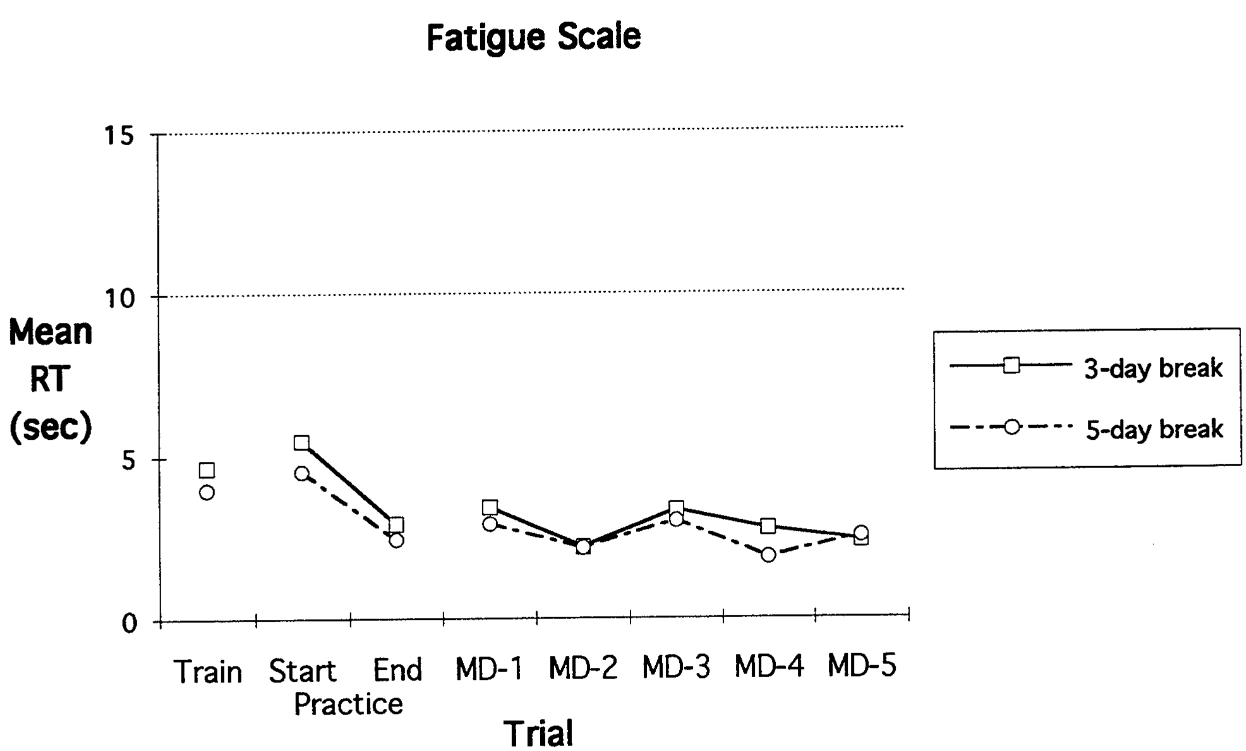


Dual - Memory Search



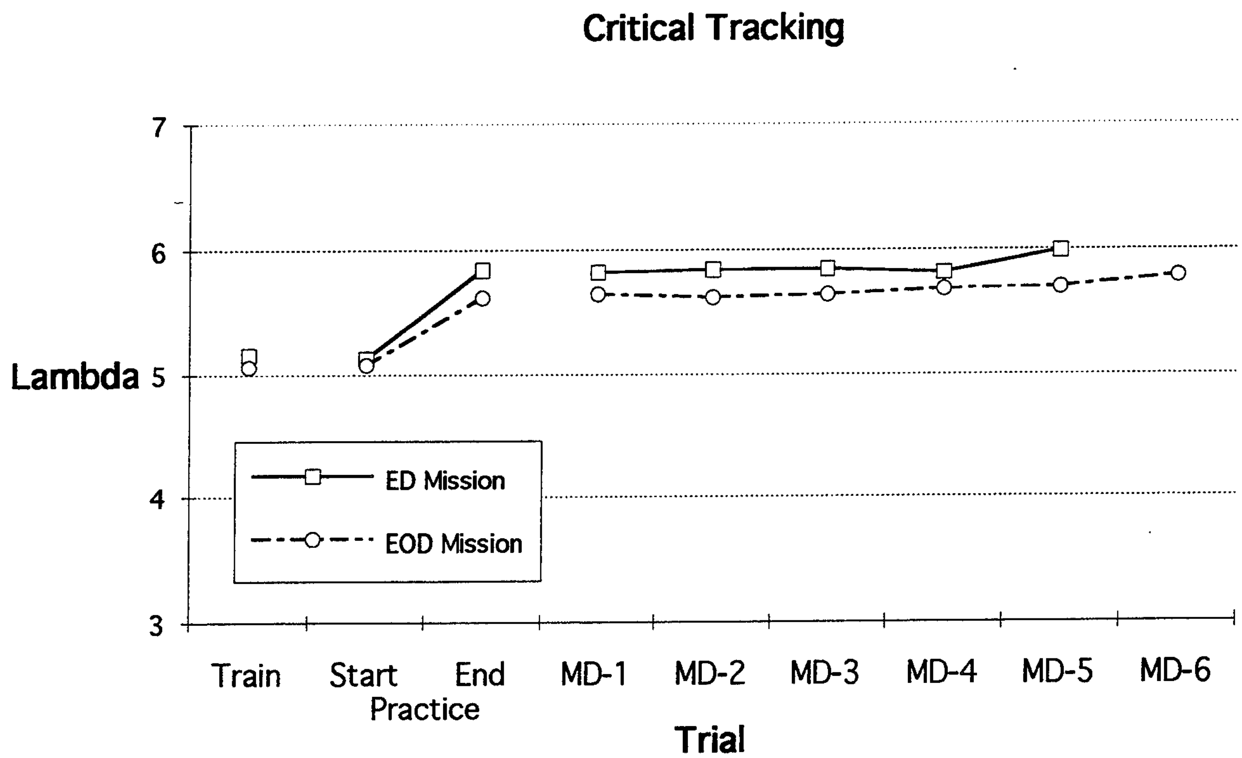

Dual - Memory Search
Throughput
120
100
80
60
40
20
0
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

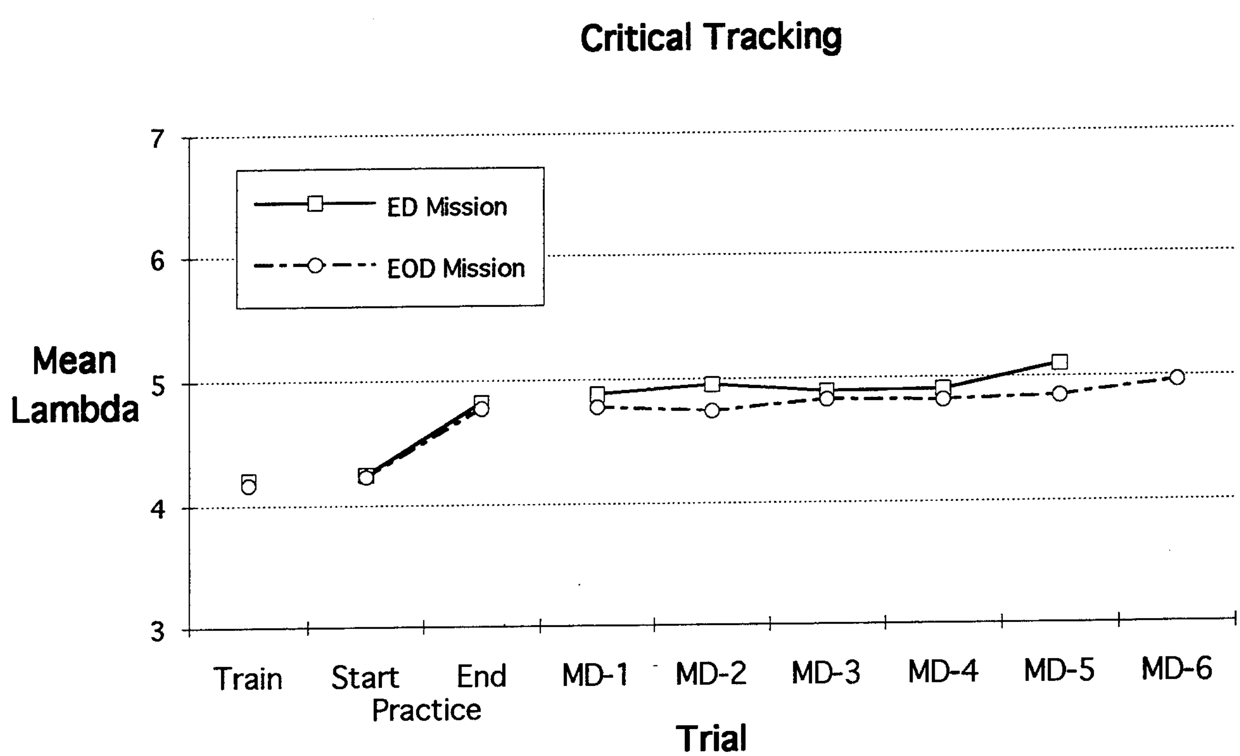

Fatigue Scale
7
6
5
4
3
2
1
0
Response
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

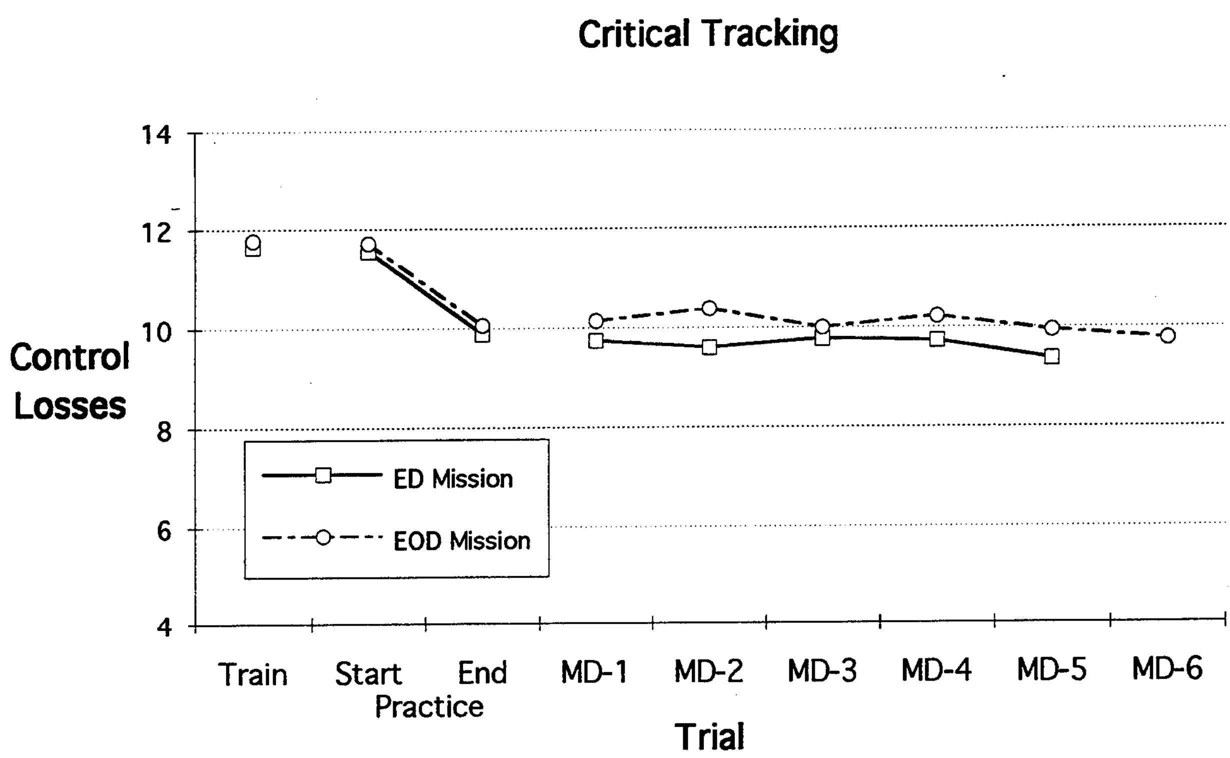

Fatigue Scale
15
10
5
0
Mean RT (sec)
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
Trial
3-day break
5-day break

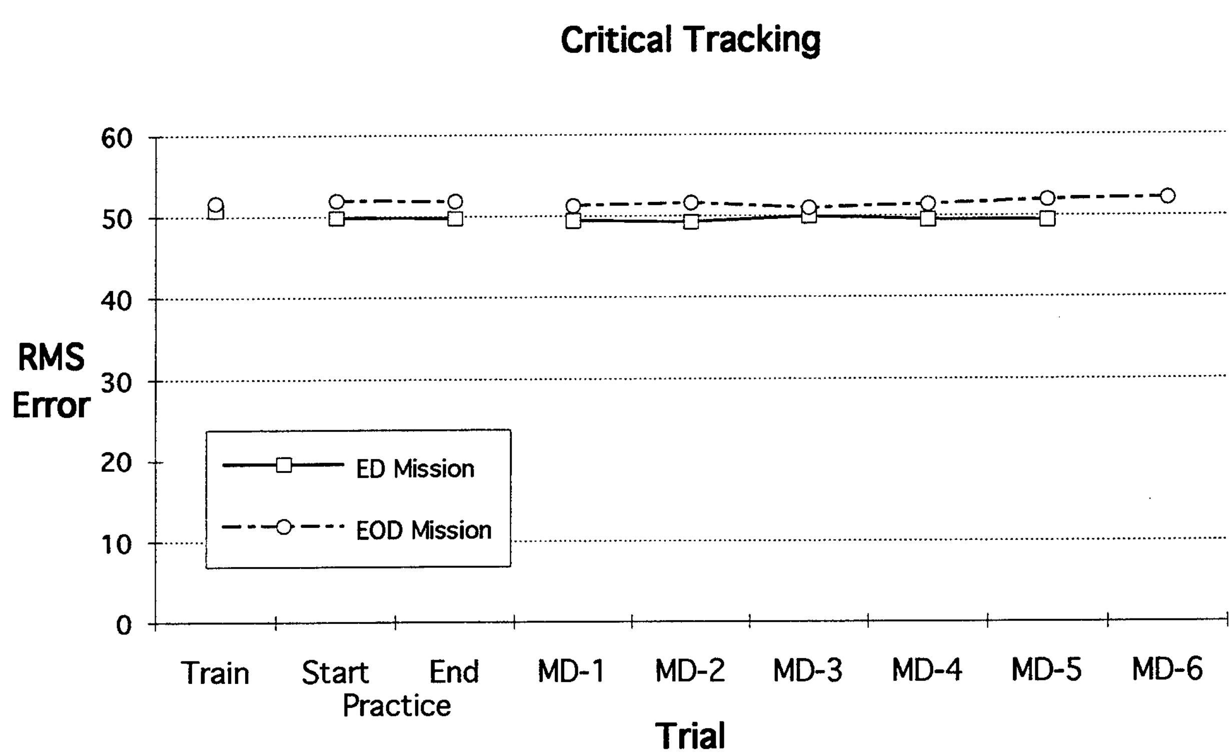
APPENDIX H

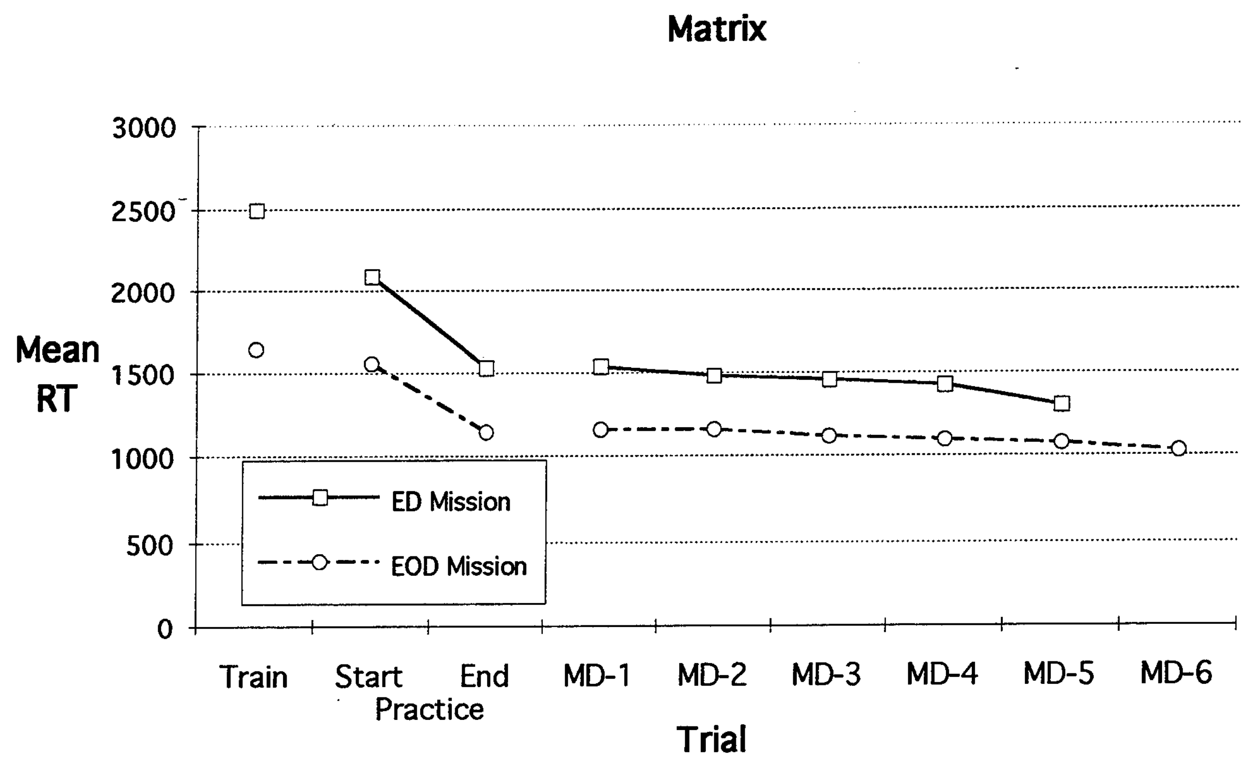
MISSION SCHEDULE DATA

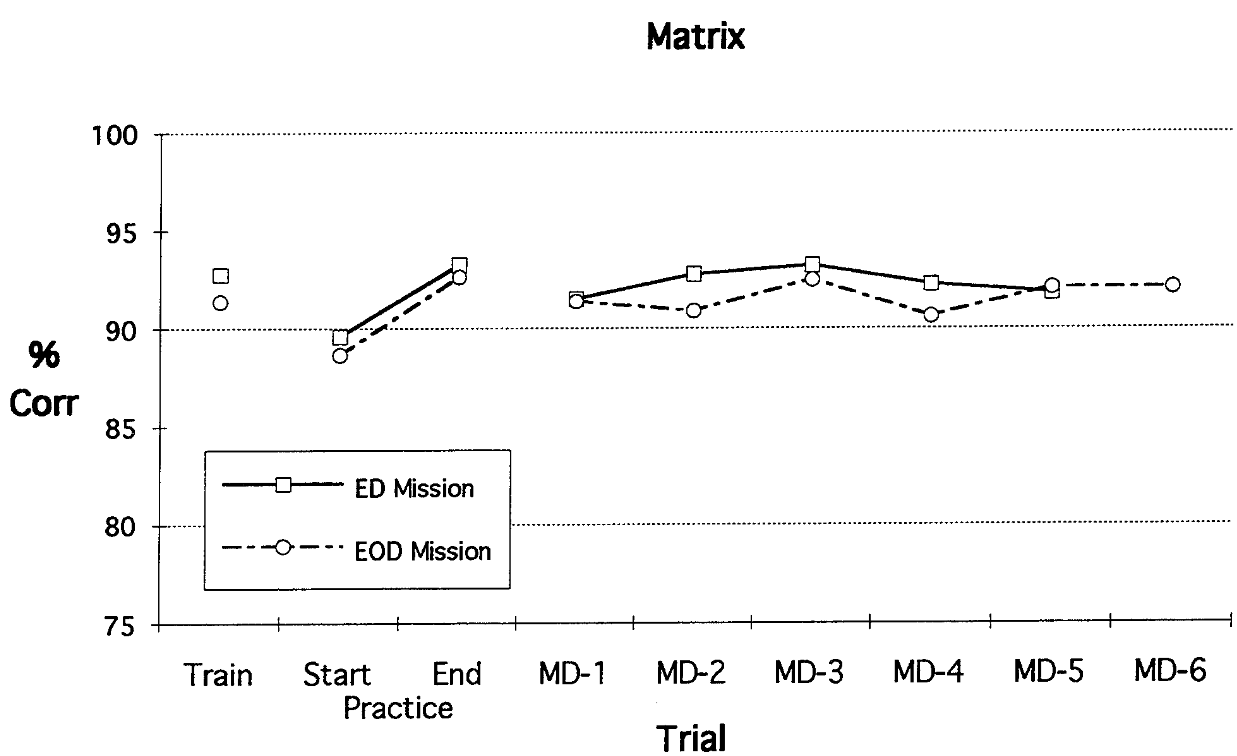

Critical Tracking
7
6
Lambda 5
4
3
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial

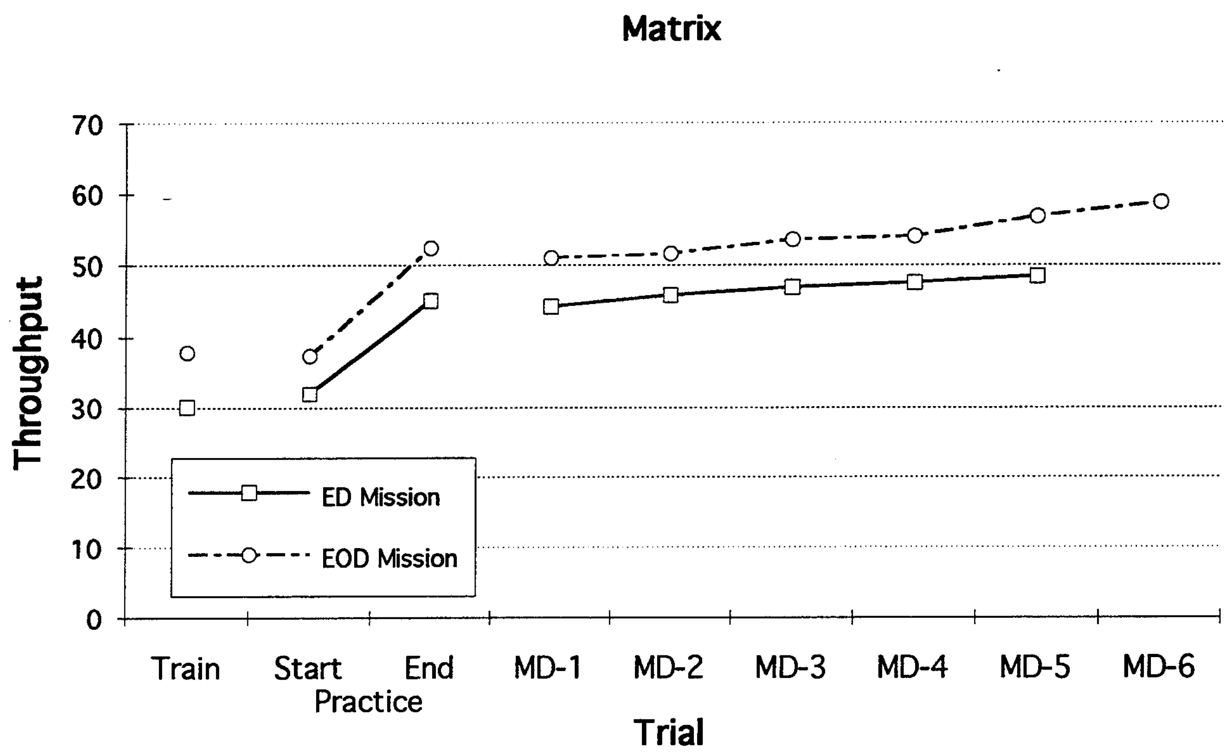

Critical Tracking
7
6
Mean Lambda 5
4
3
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial


Critical Tracking
14
12
10
8
6
4
Control Losses
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial

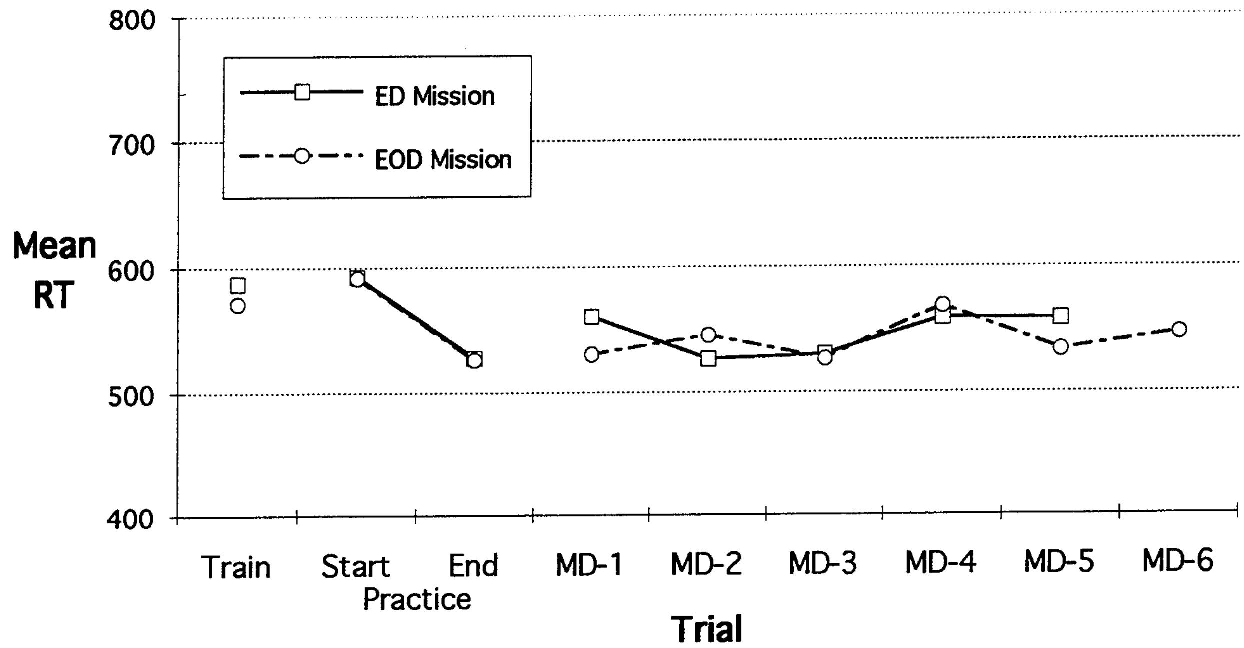

Critical Tracking
60
50
40
30
20
10
0
RMS Error
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial


Matrix
3000
2500
2000
1500
1000
500
0
Mean
RT
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial

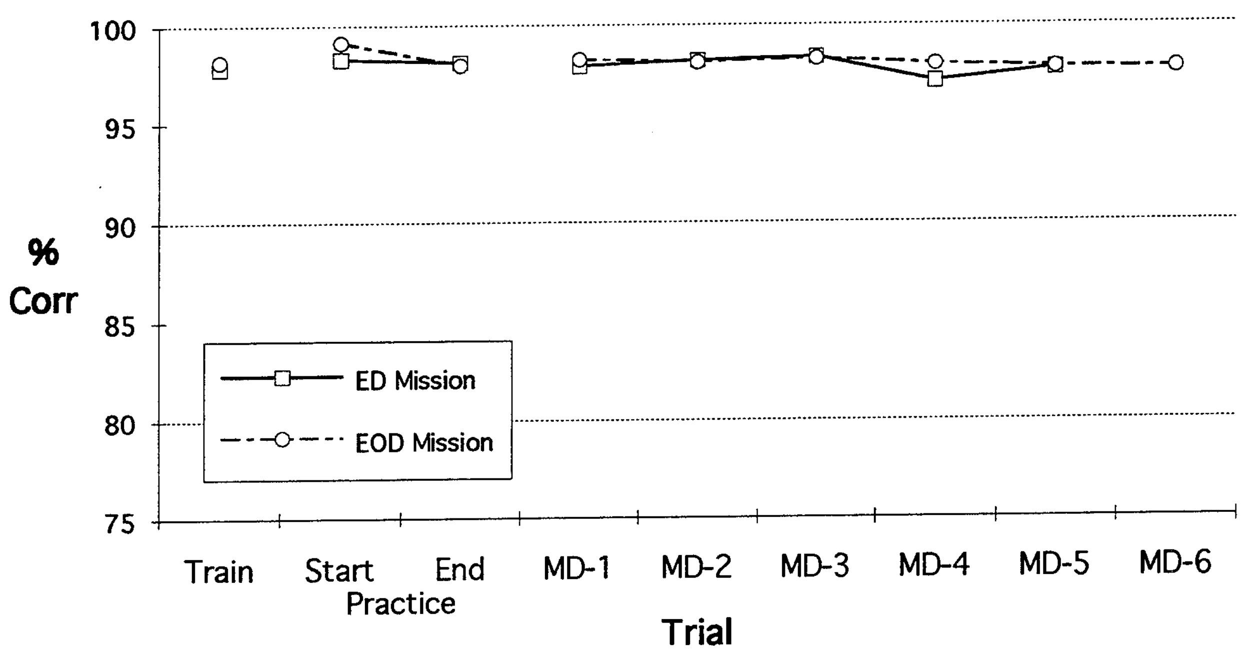

Matrix
100
95
90
85
80
75
%
Corr
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial

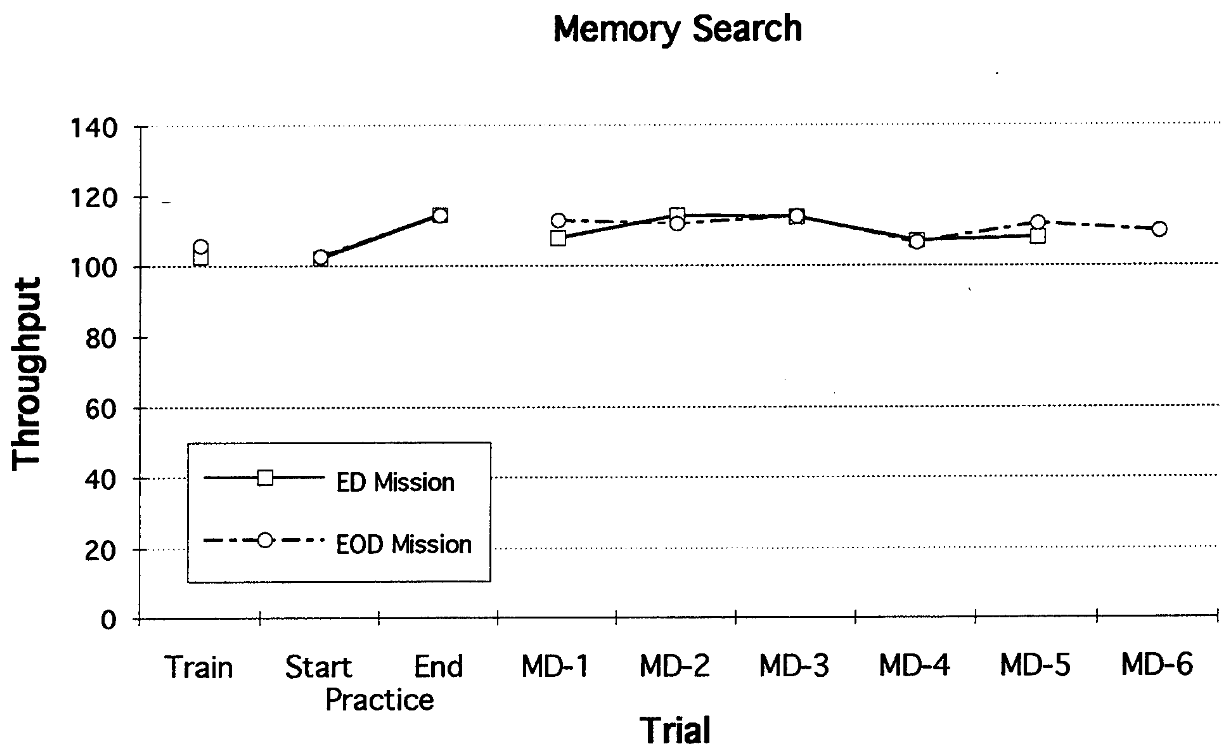

Matrix
70
60
50
40
30
20
10
0
Throughput
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial

Memory Search



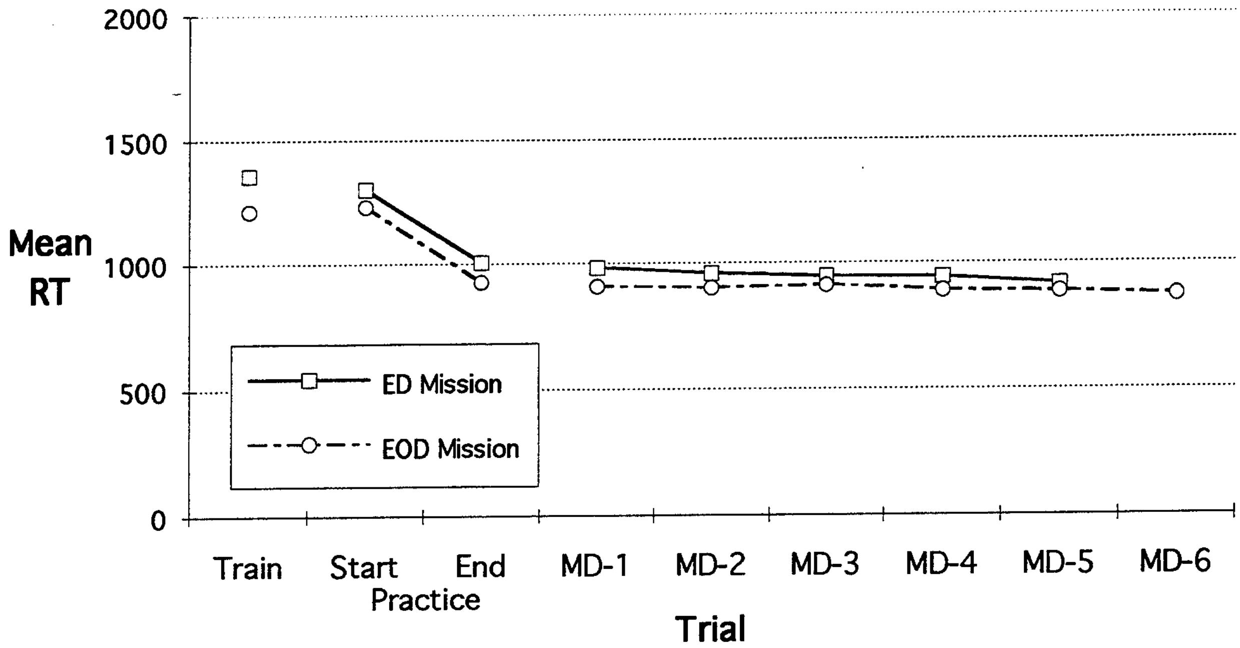
Memory Search



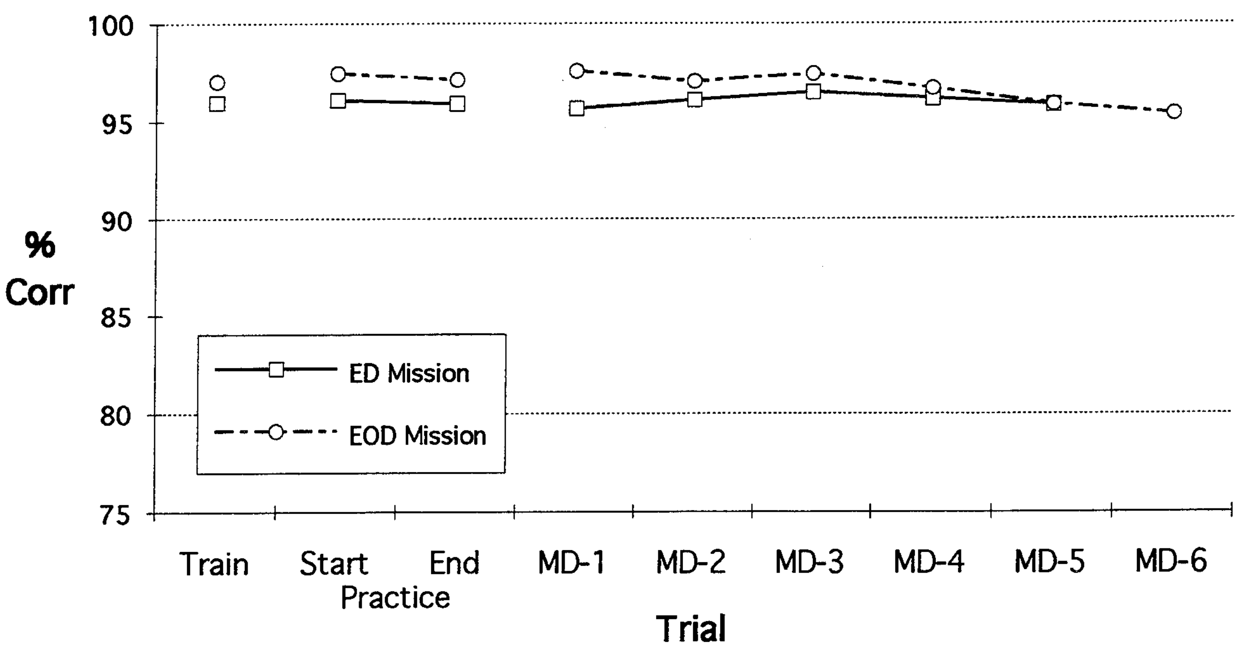


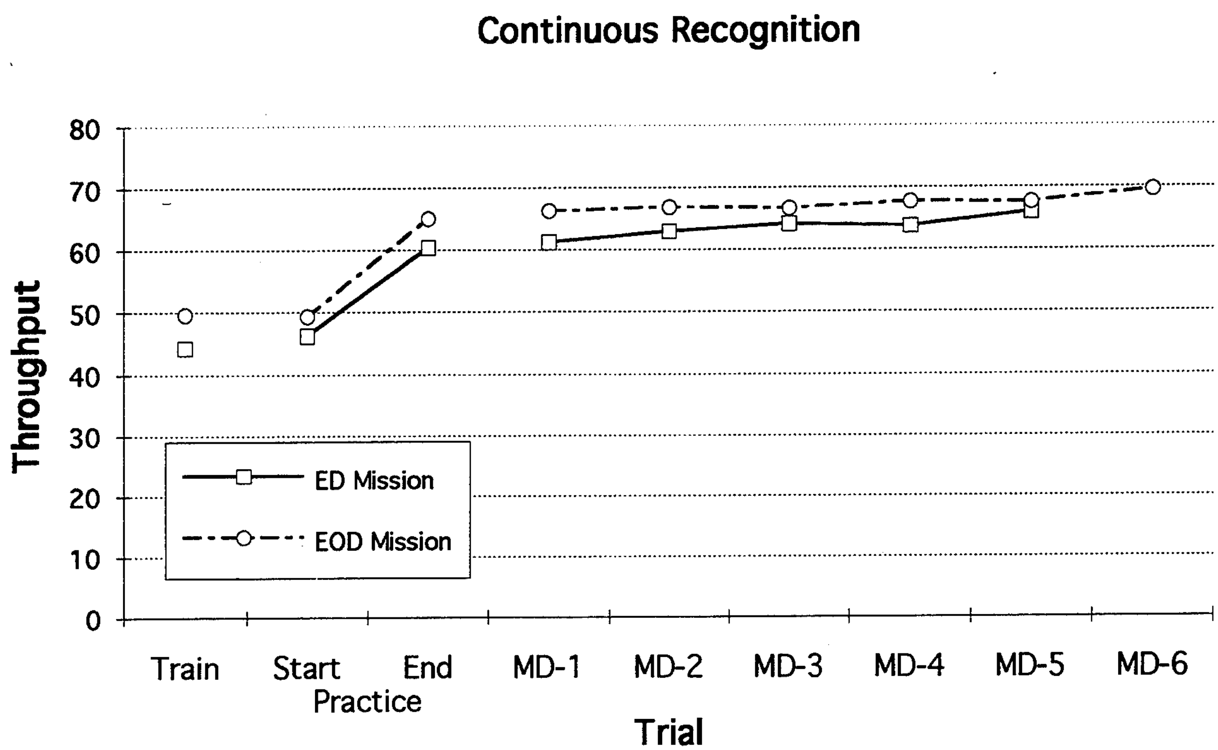
Memory Search
Throughput
140
120
100
80
60
40
20
0
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial

Continuous Recognition



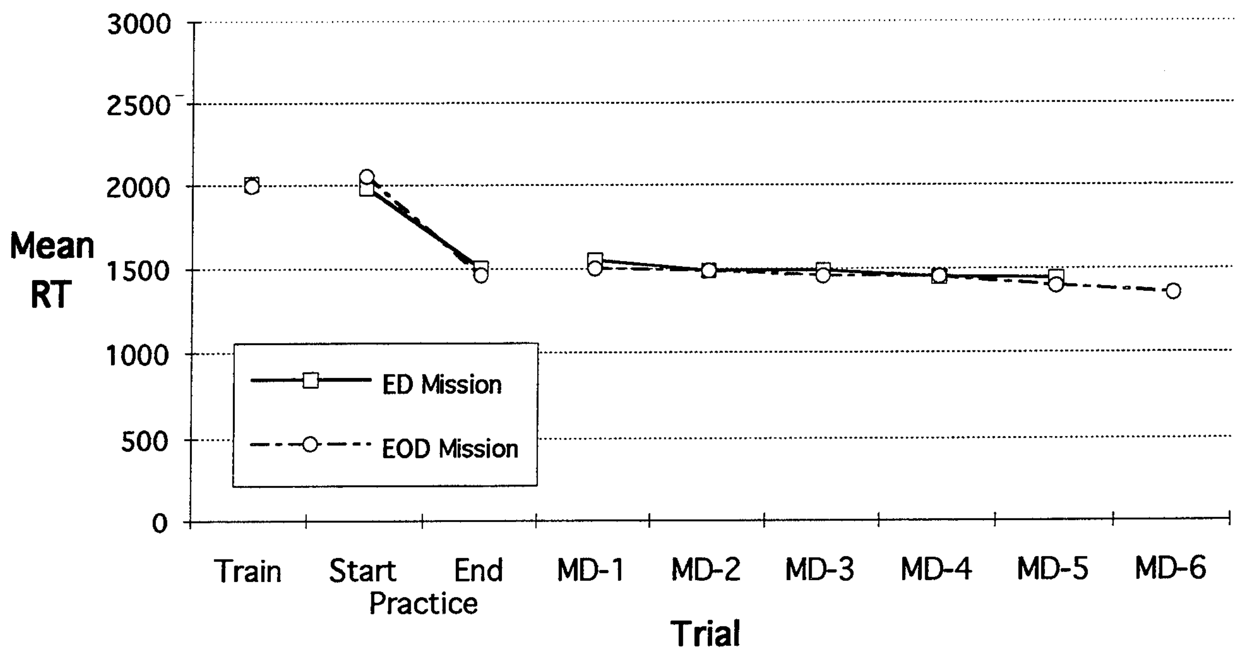
Continuous Recognition



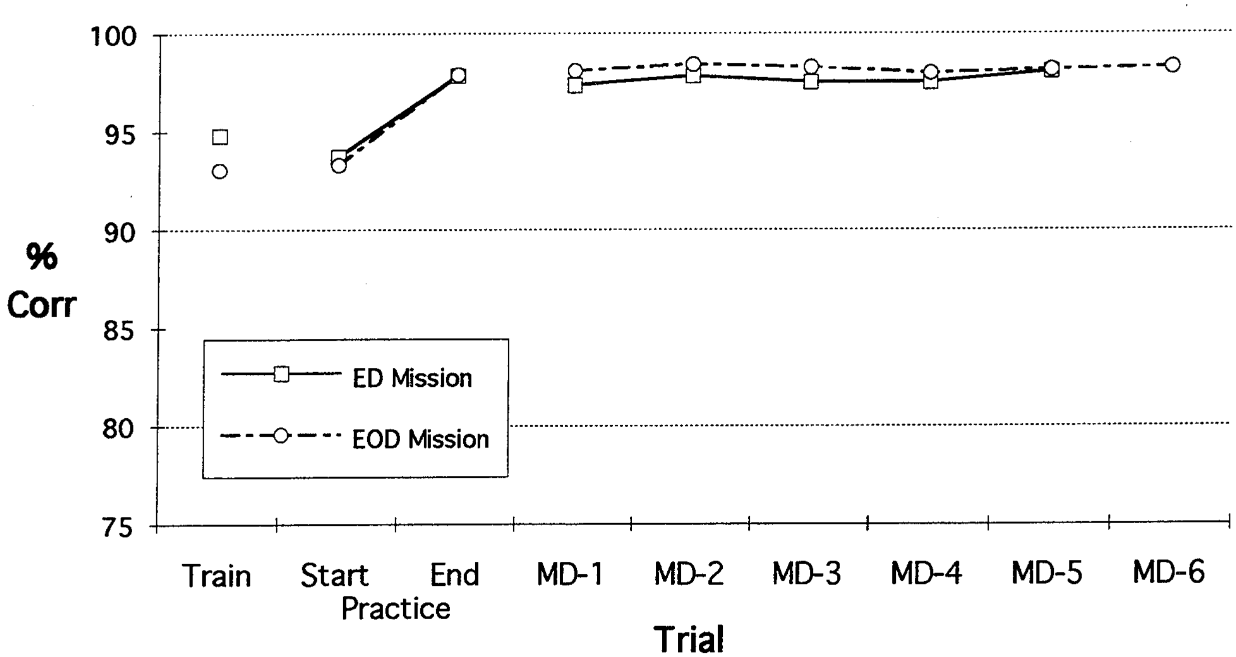


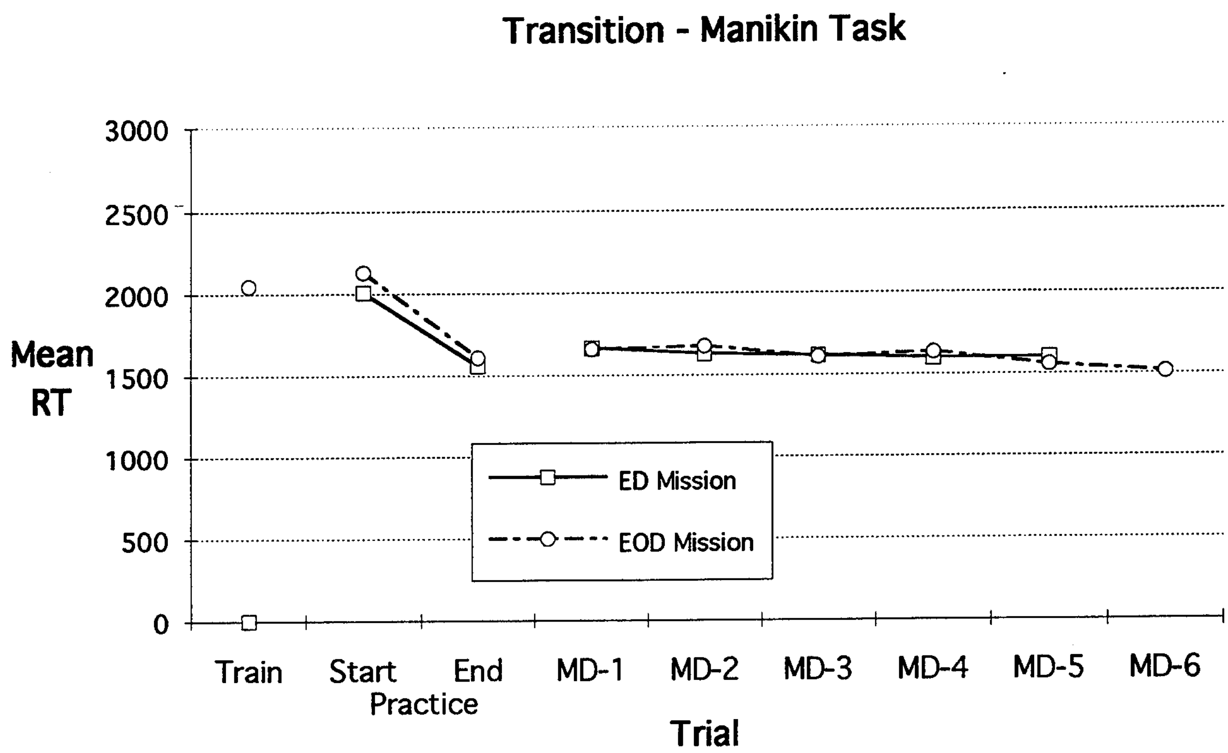
Continuous Recognition
Throughput
80
70
60
50
40
30
20
10
0
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial

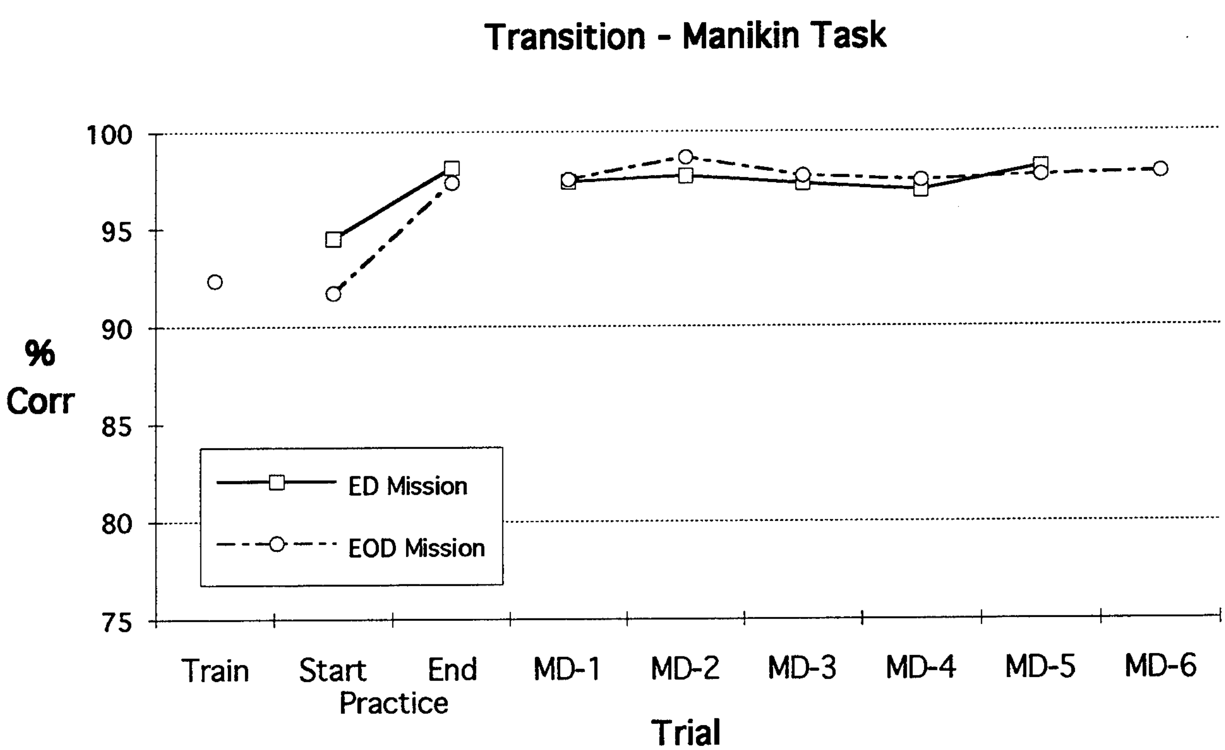
Switching - Manikin Task



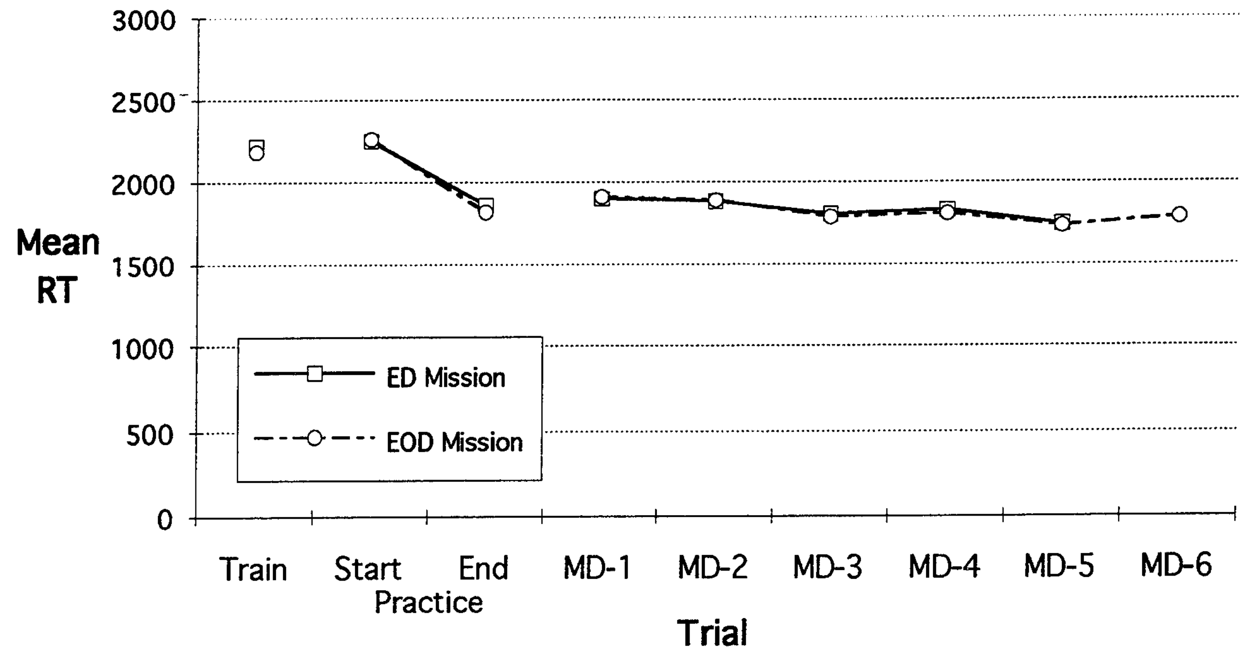
Switching - Manikin Task



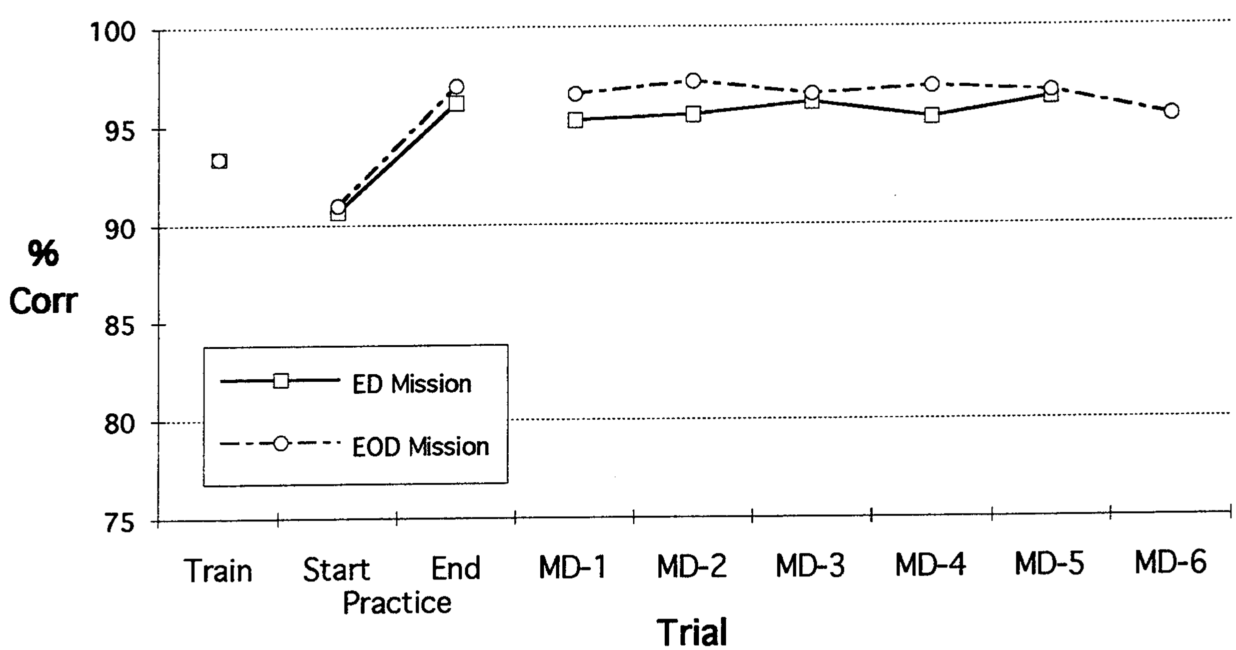

Transition - Manikin Task
3000
2500
2000
1500
1000
500
0
Mean
RT
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial

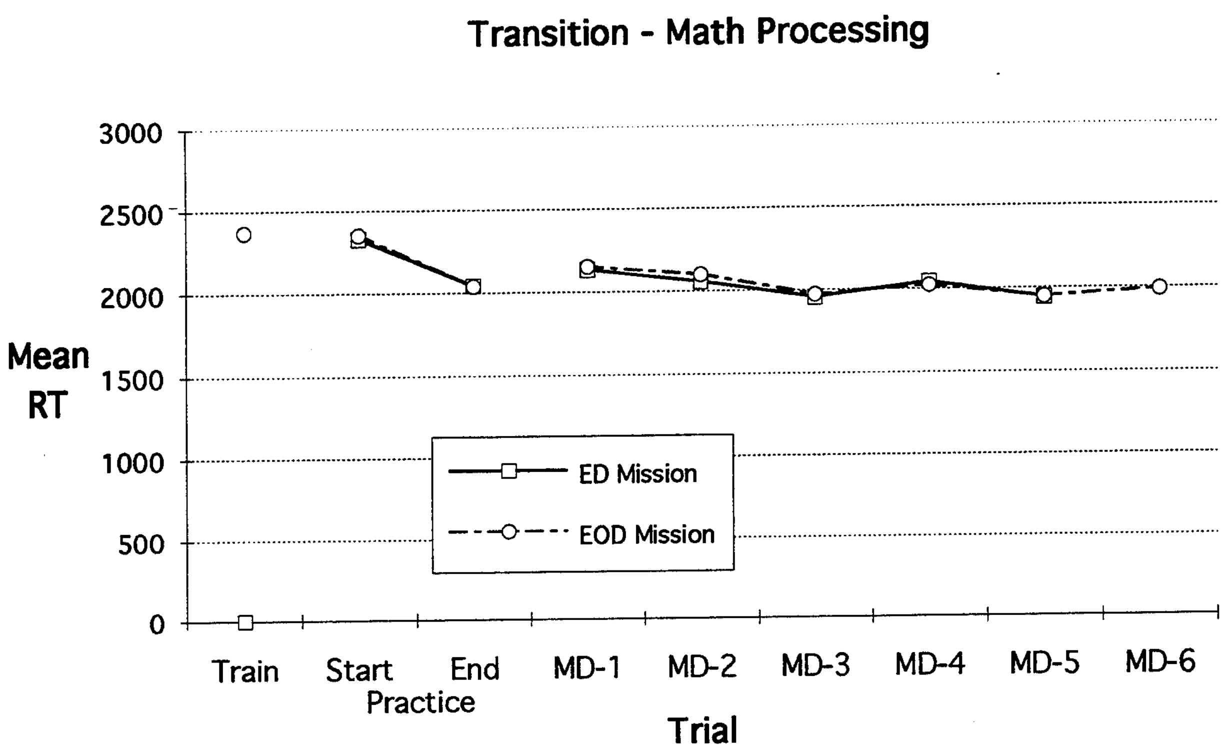

Transition - Manikin Task
100
95
90
85
80
75
%
Corr
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial

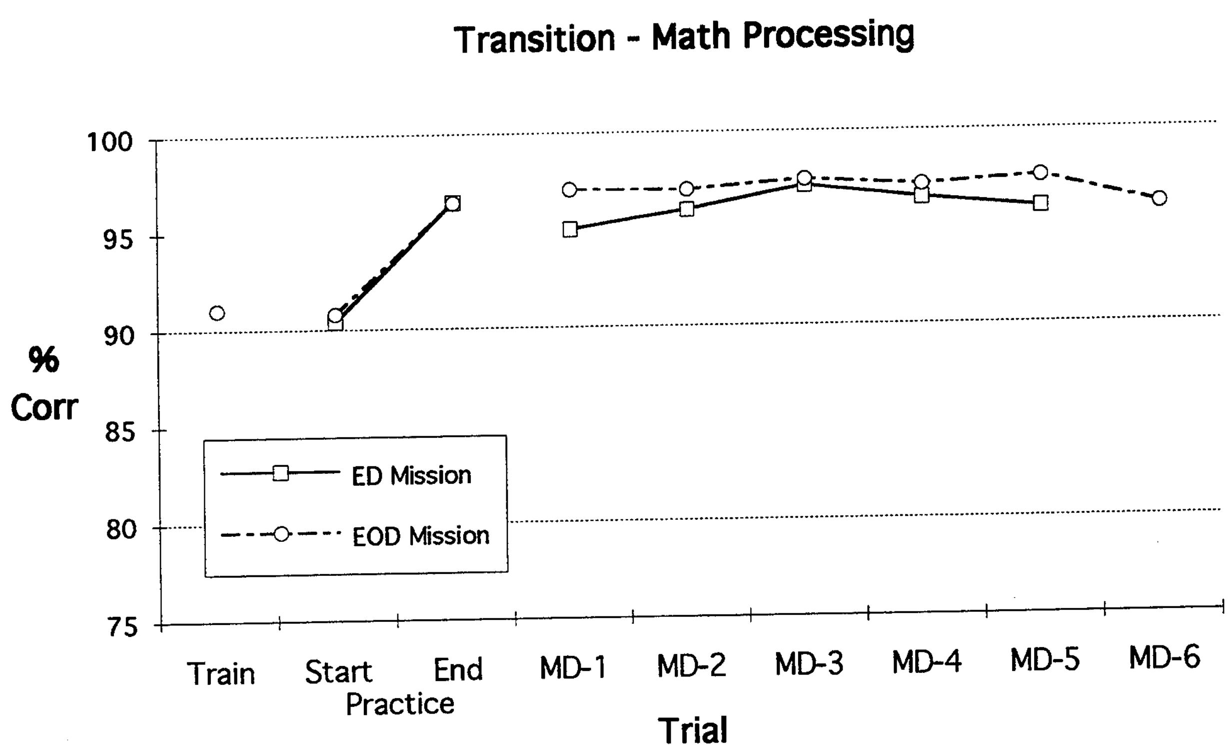
Switching - Math Processing

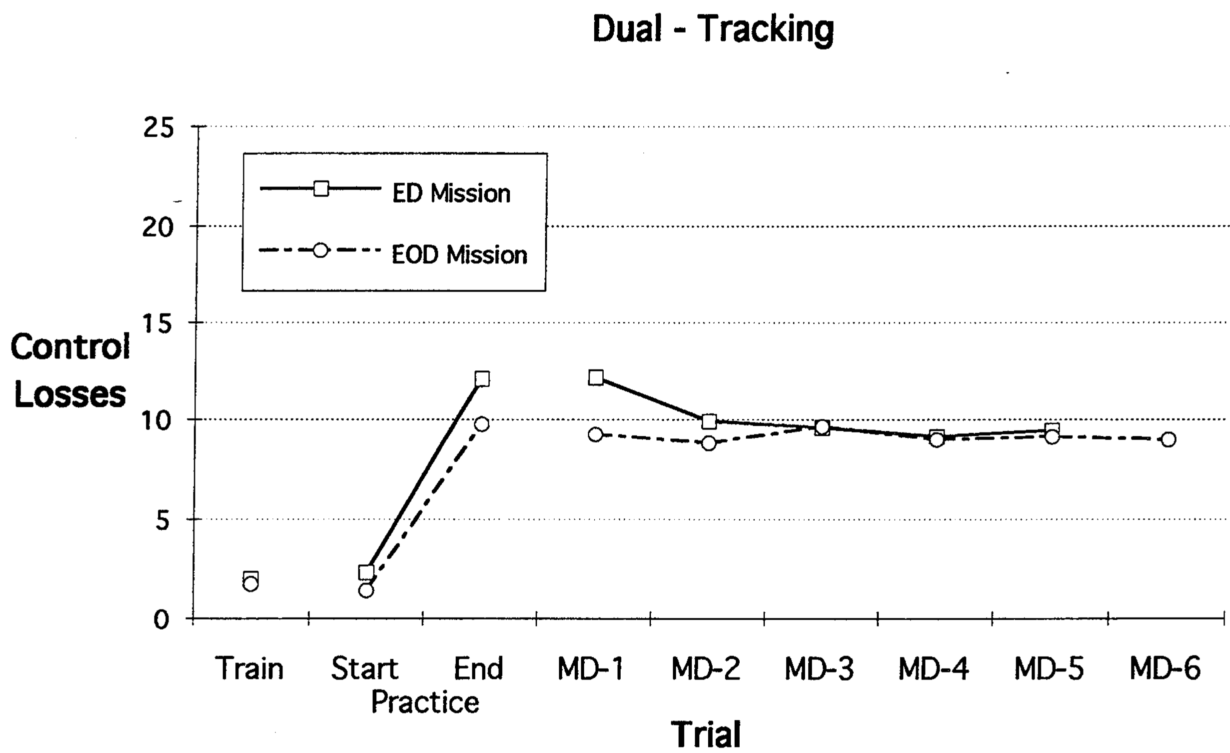


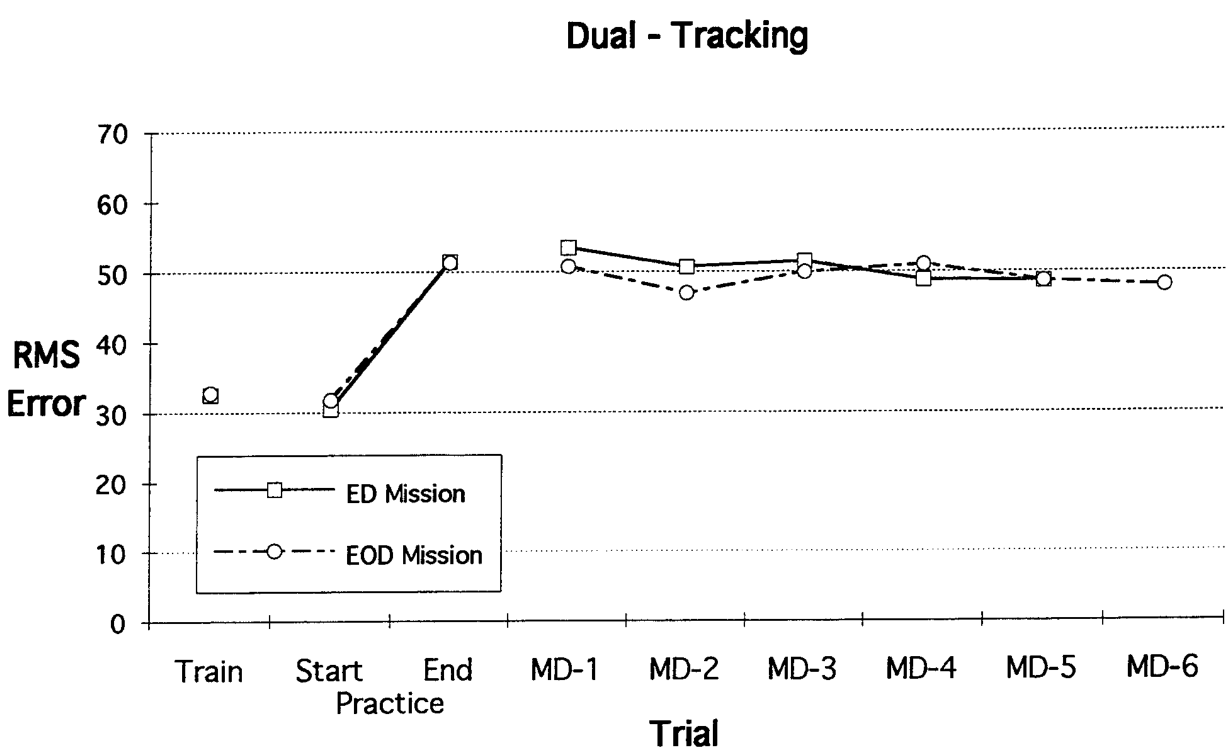
Switching - Math Processing



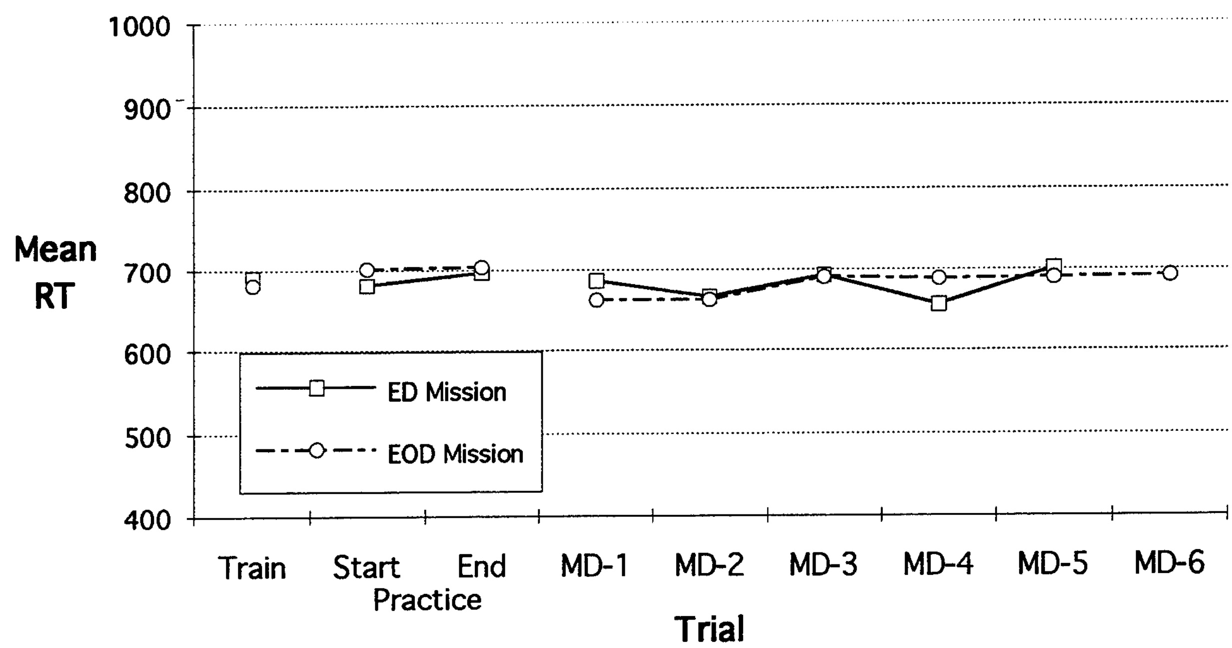

Transition - Math Processing
3000
2500
2000
1500
1000
500
0
Mean RT
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial


Transition - Math Processing
100
95
90
85
80
75
% Corr
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial

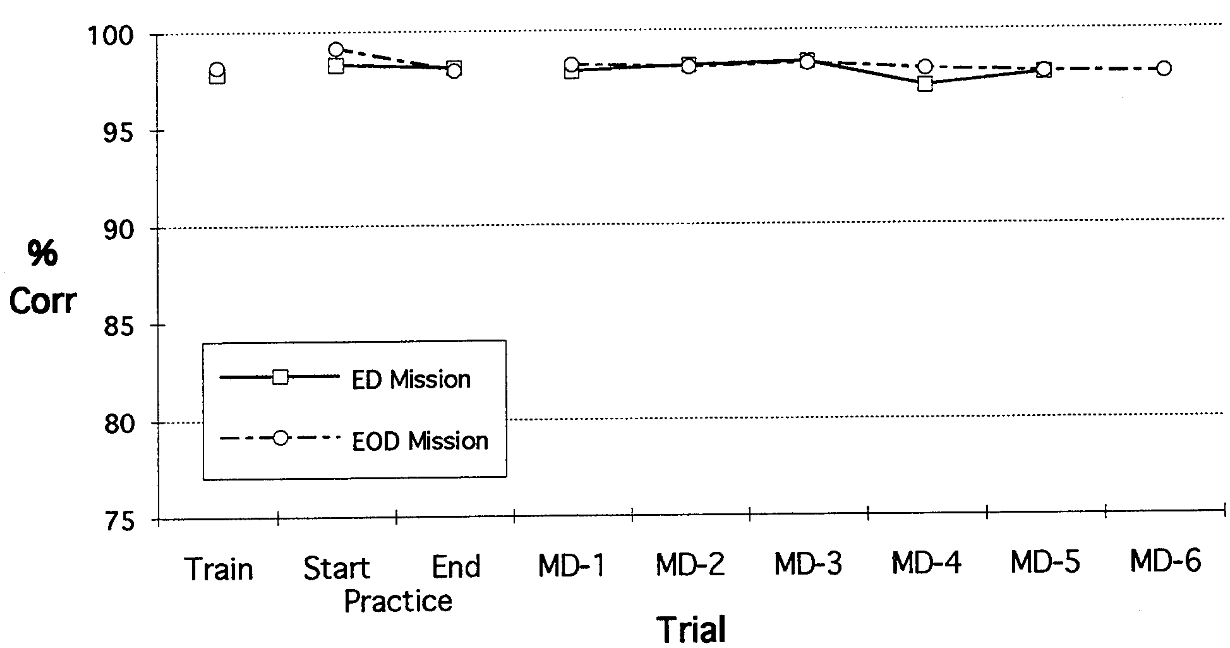

Dual - Tracking
25
20
15
10
5
0
Control Losses
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial

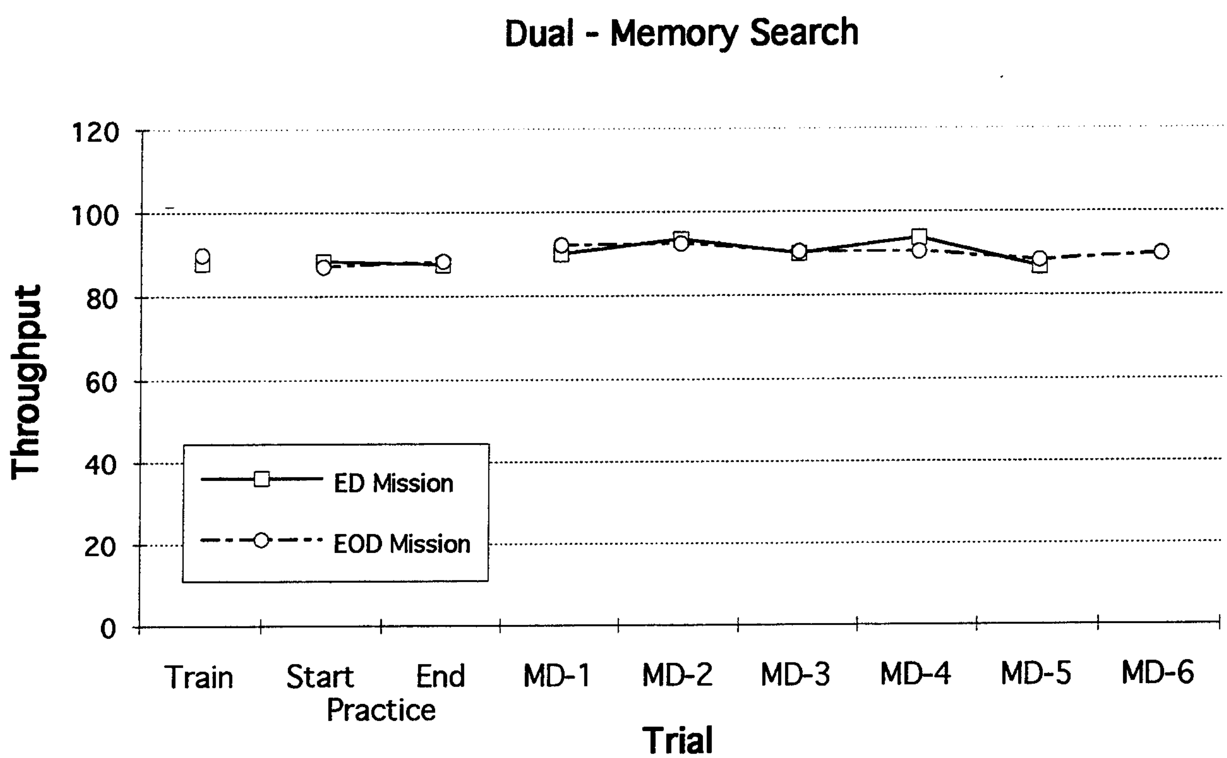

Dual - Tracking
70
60
50
40
30
20
10
0
RMS Error
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial

Dual - Memory Search

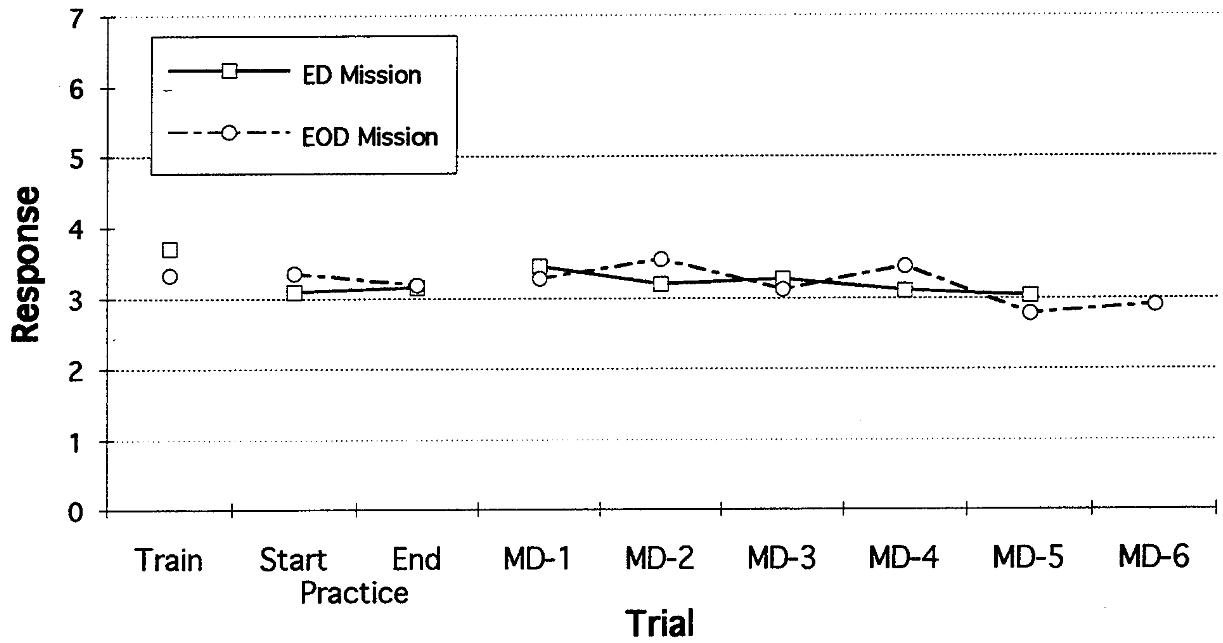


Memory Search

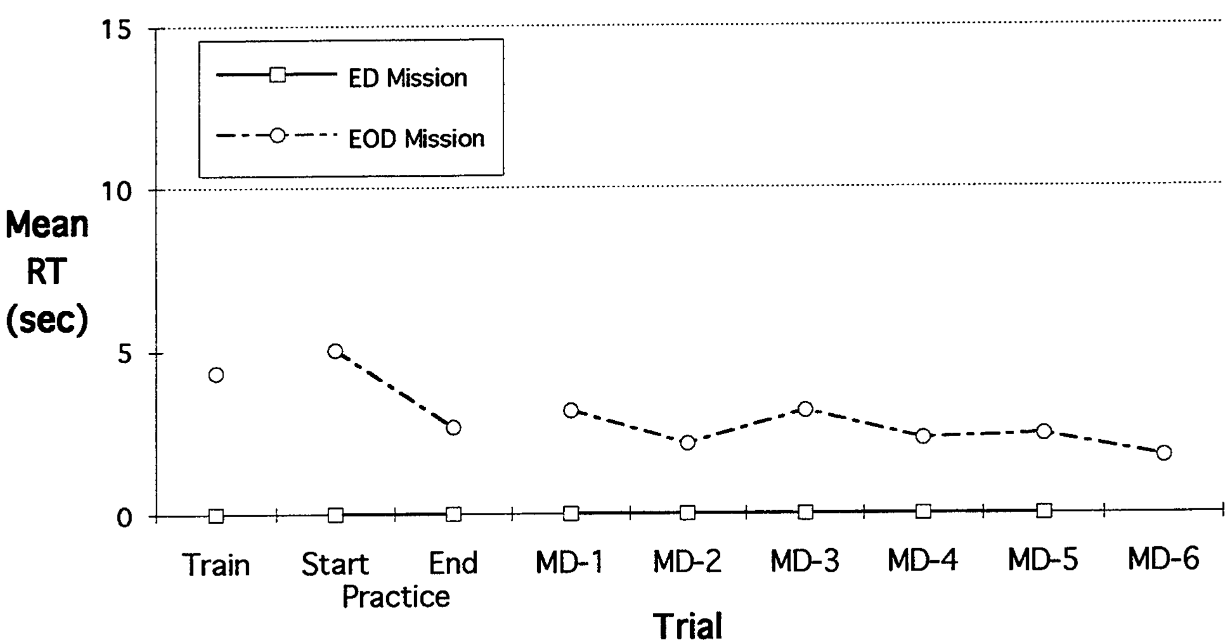



Dual - Memory Search
Throughput
120
100
80
60
40
20
0
ED Mission
EOD Mission
Train
Start
End
Practice
MD-1
MD-2
MD-3
MD-4
MD-5
MD-6
Trial

Fatigue Scale

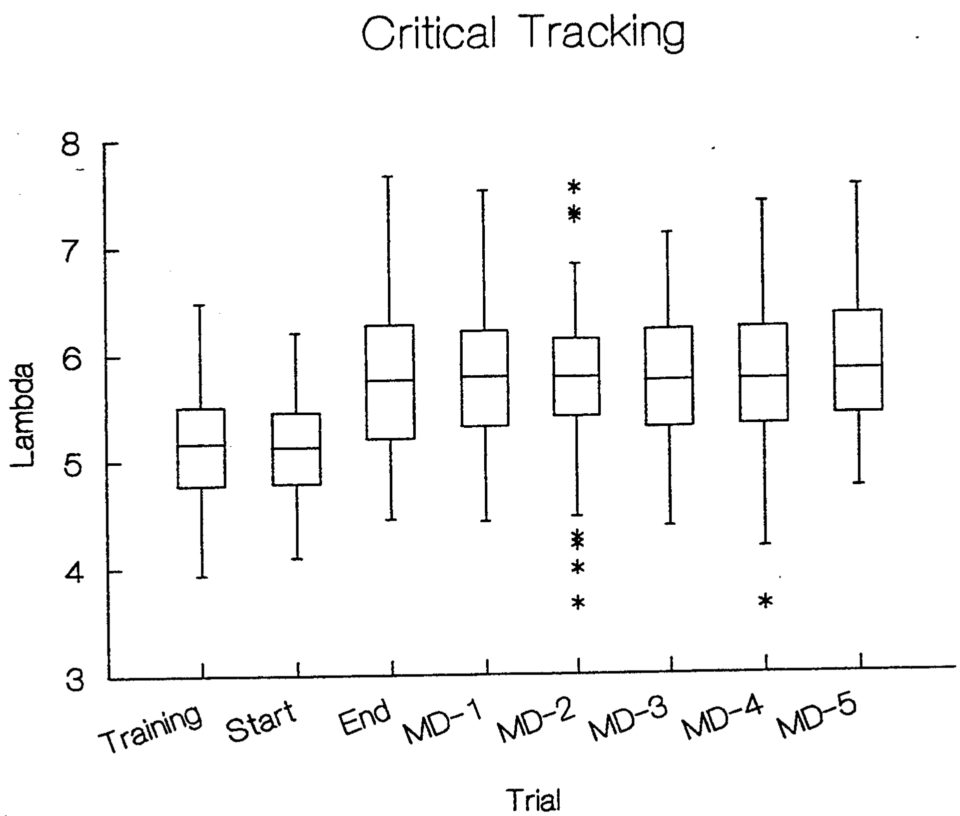


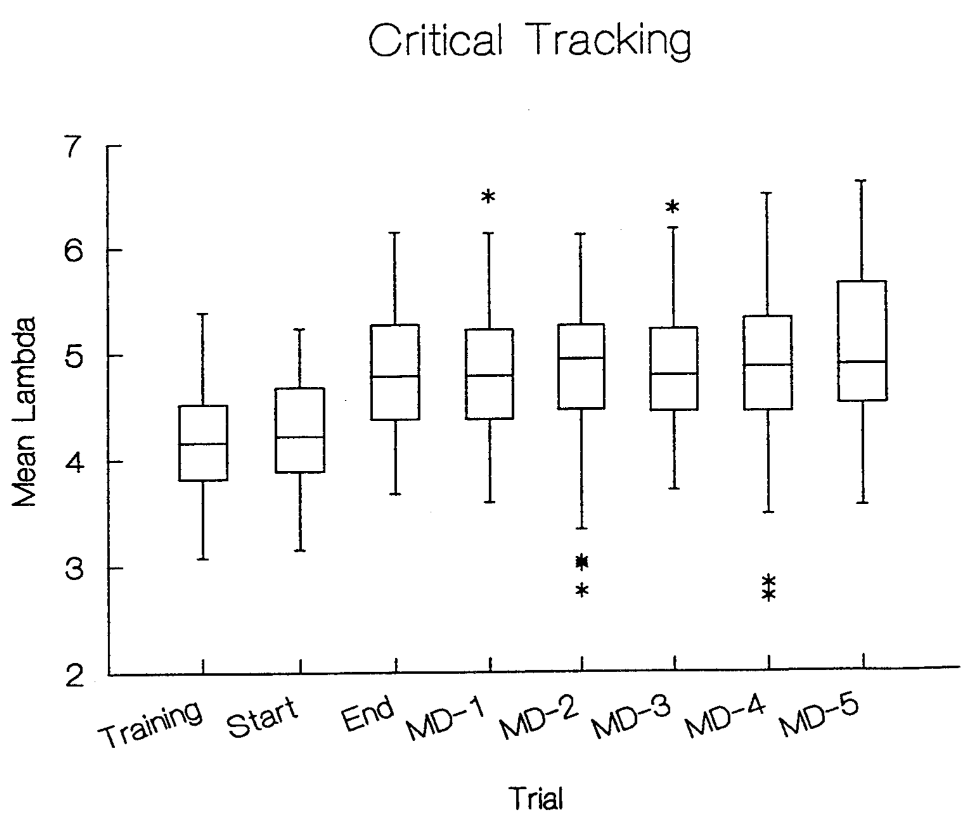
Fatigue Scale

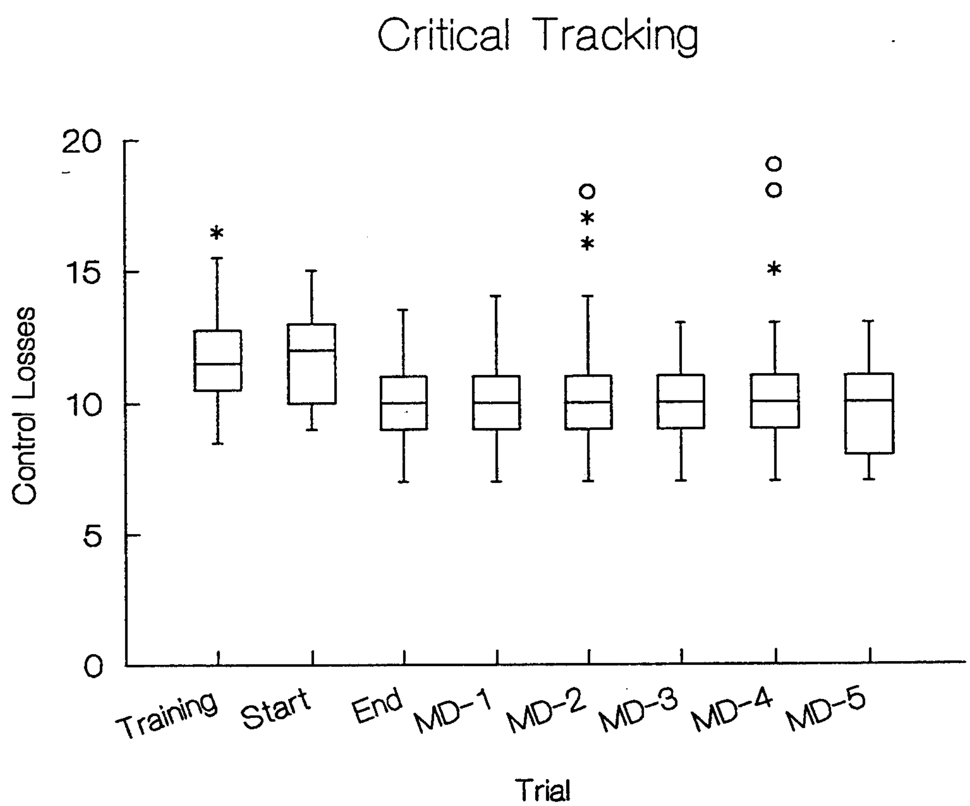


APPENDIX I

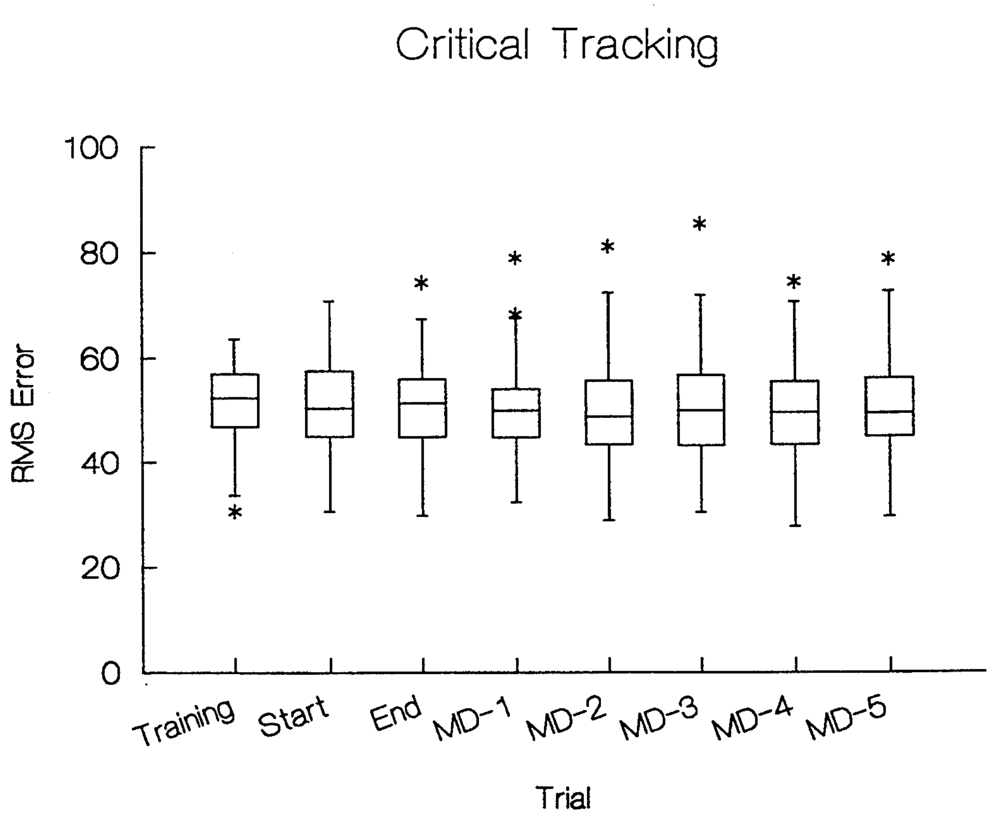
BOX-AND-WHISKER PLOTS (PERFORMANCE DATA)


Critical Tracking
Lambda
8
7
6
5
4
3
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

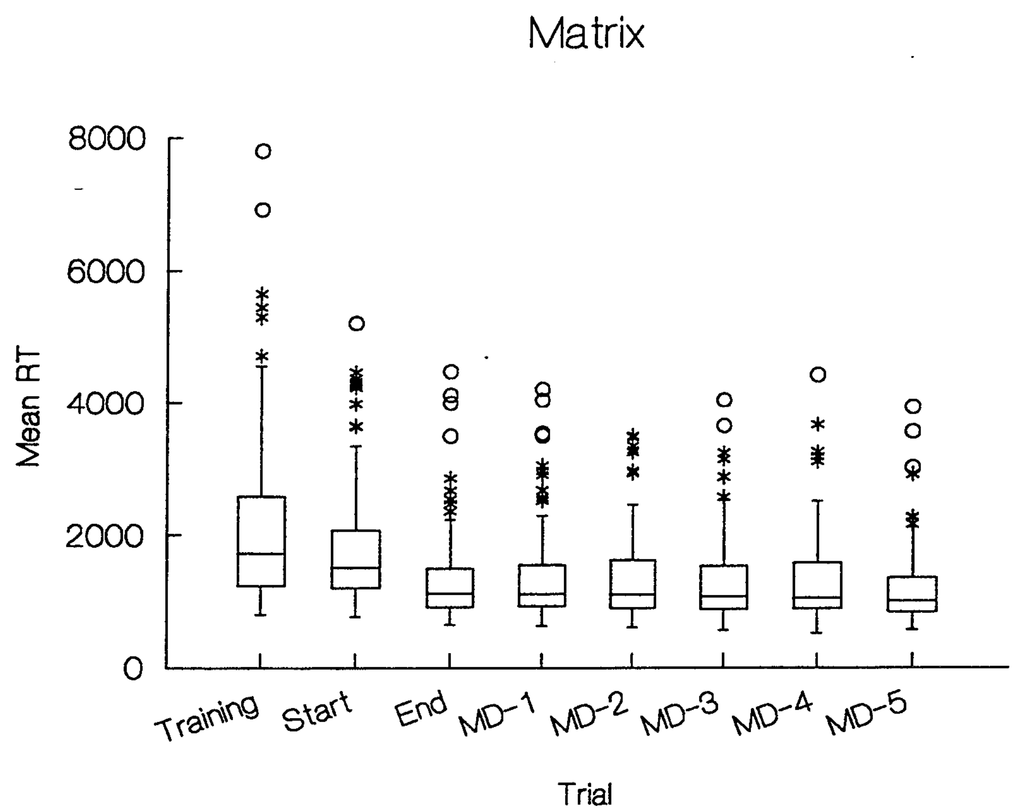

Critical Tracking
Mean Lambda
7
6
5
4
3
2
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

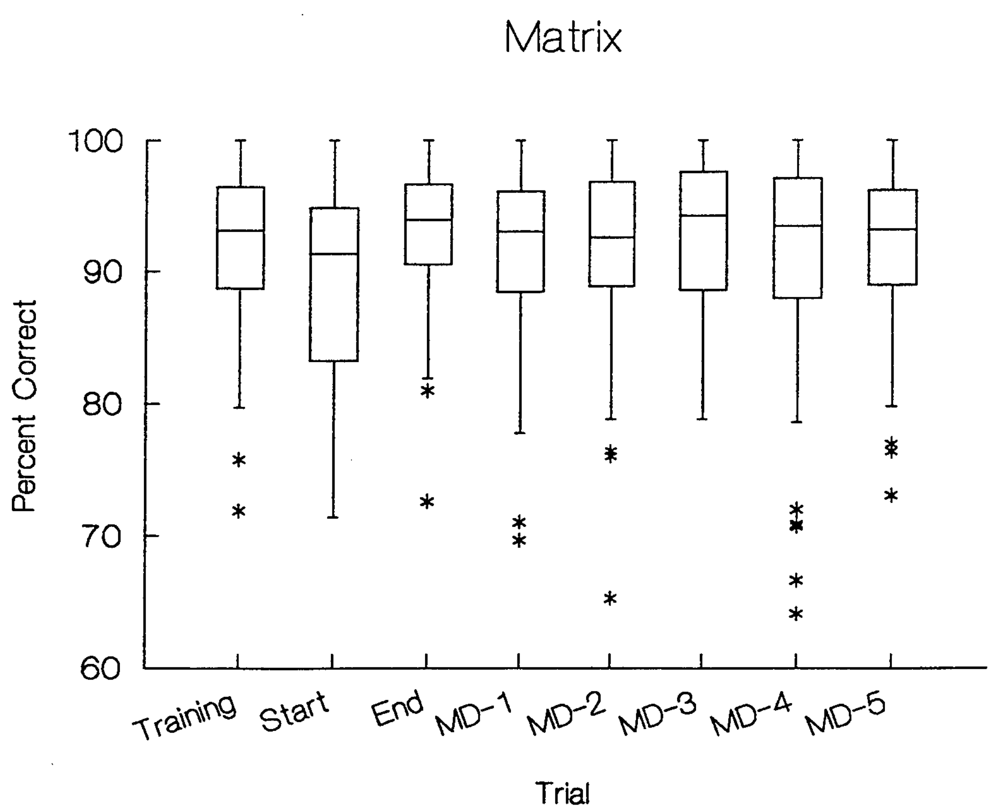


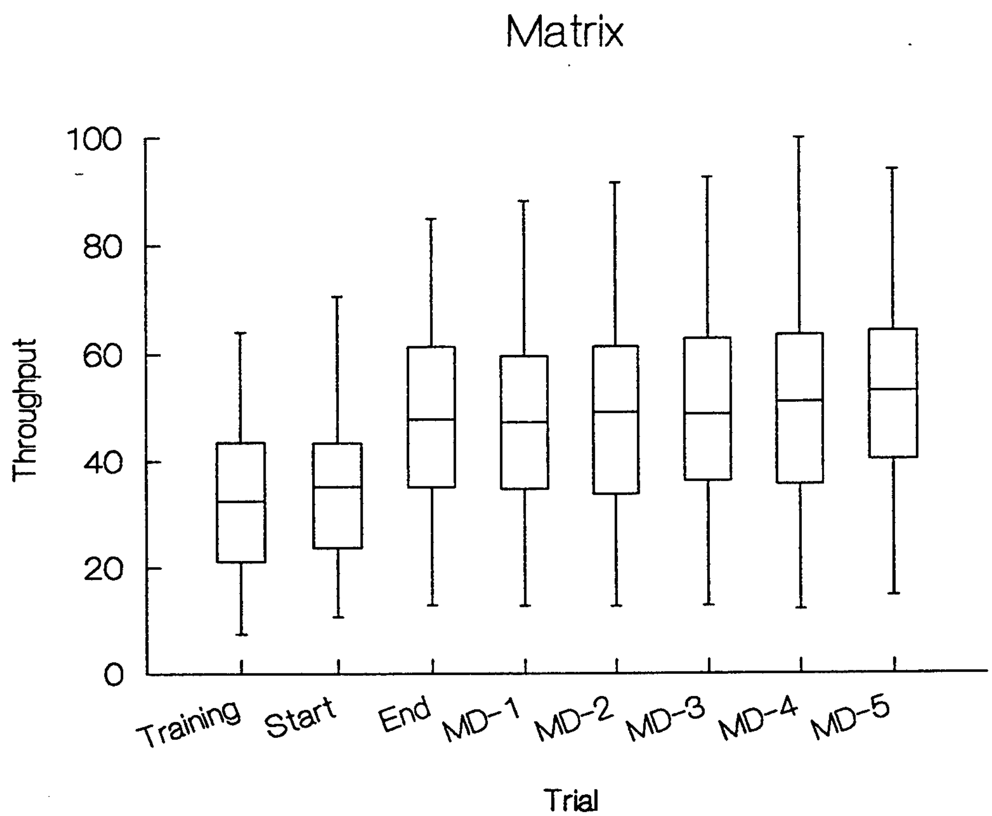
Critical Tracking
Control Losses
20
15
10
5
0
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

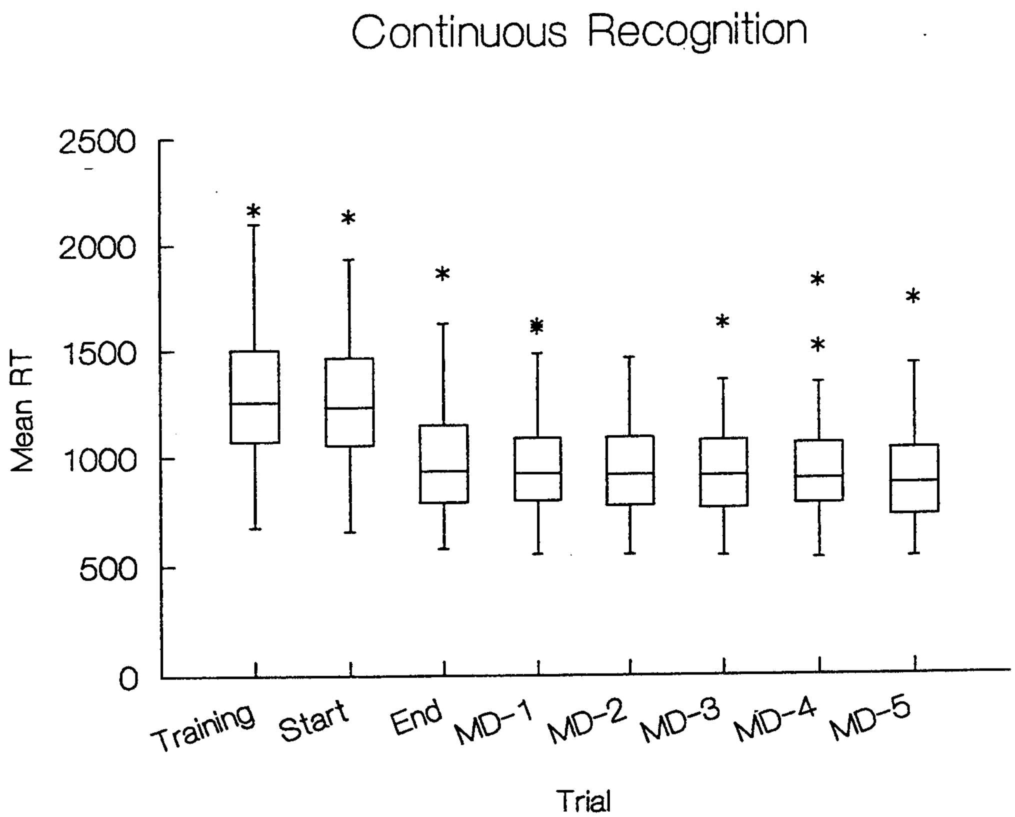


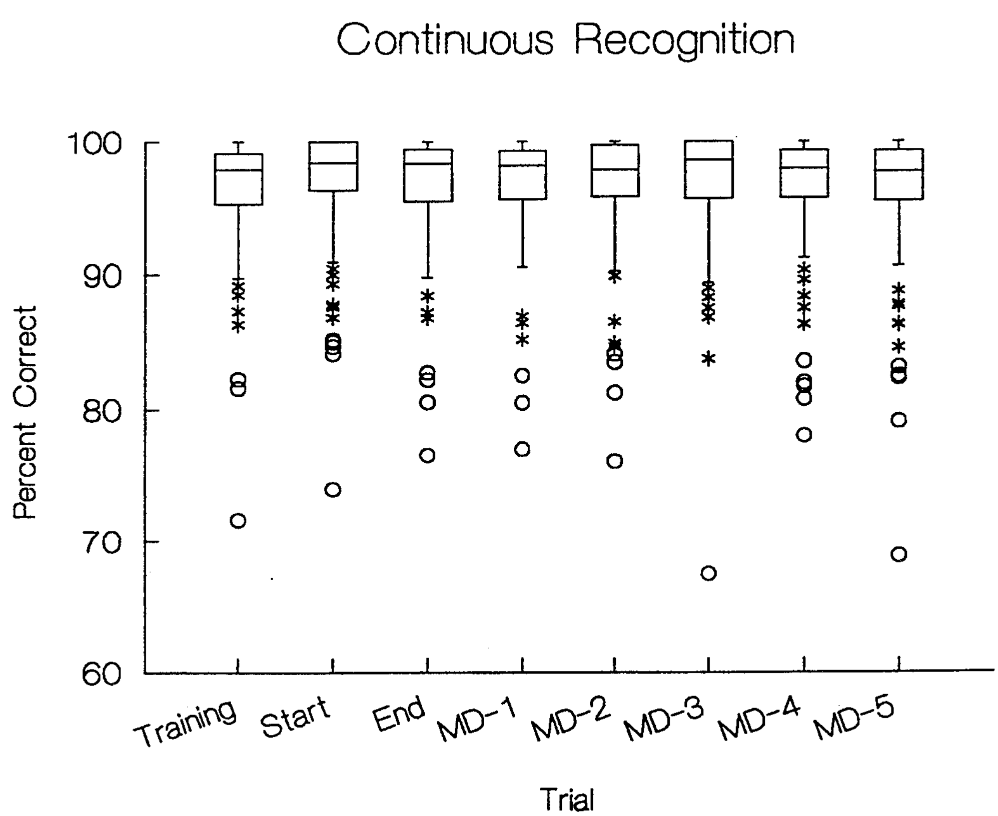
Critical Tracking
RMS Error
100
80
60
40
20
0
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

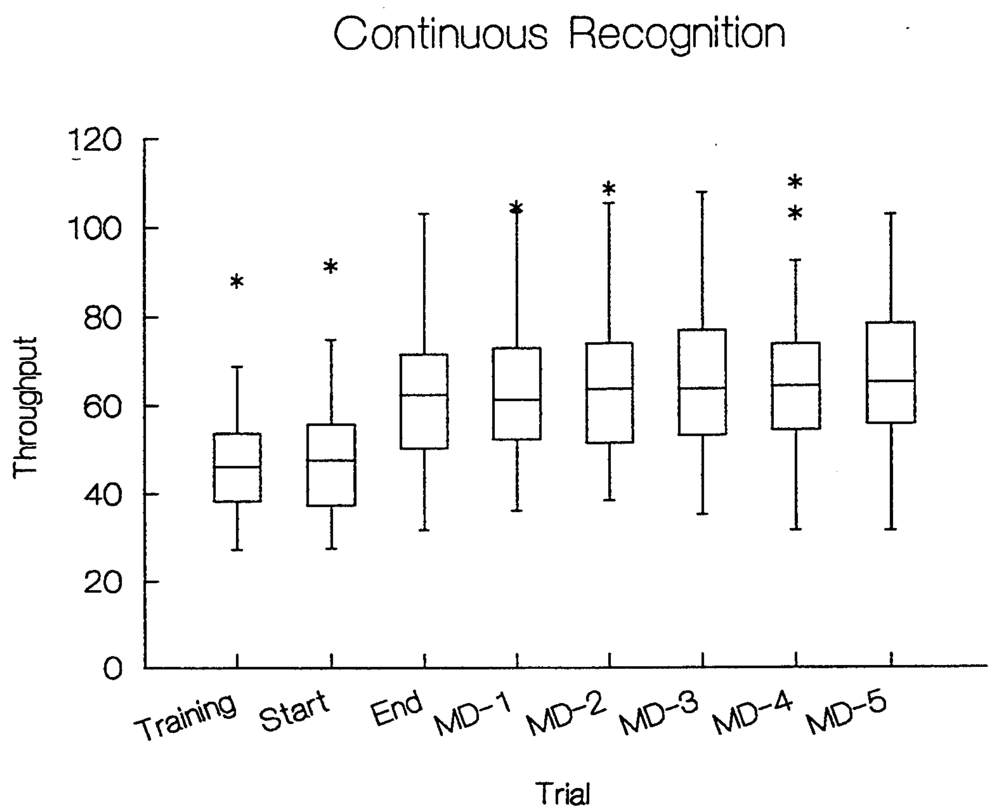

Matrix
8000
6000
4000
2000
0
Mean RT
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial


Matrix
100
90
80
70
60
Percent Correct
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

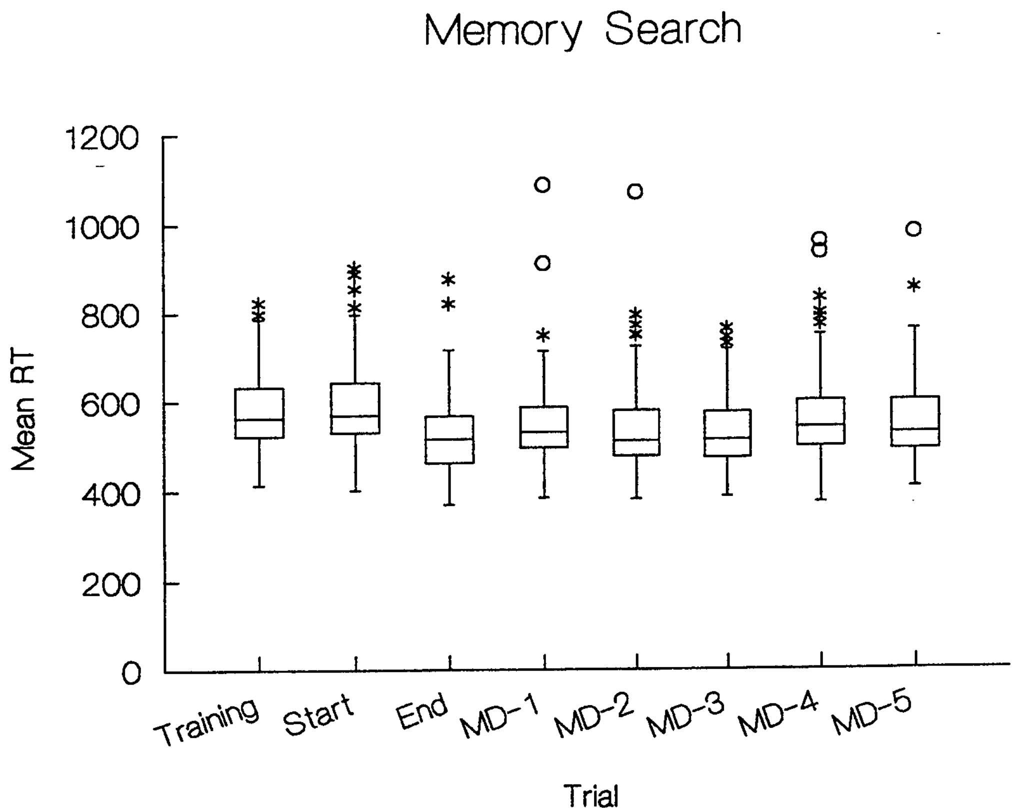

Matrix
Throughput
100
80
60
40
20
0
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

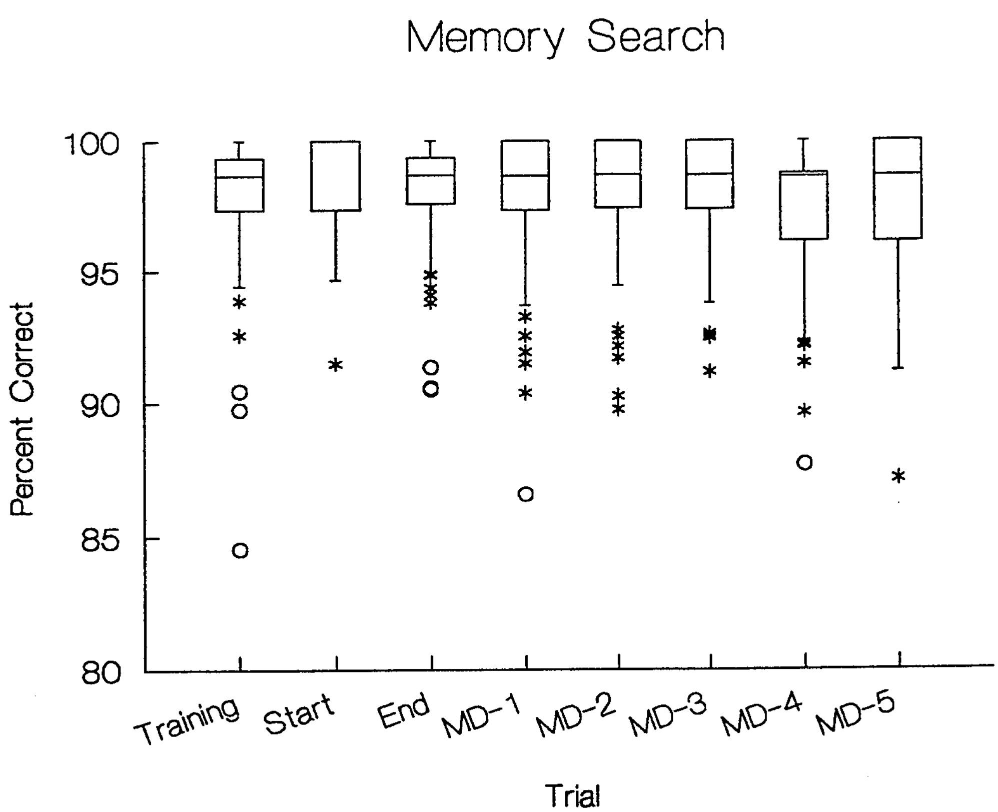

Continuous Recognition
2500
2000
1500
1000
500
0
Mean RT
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

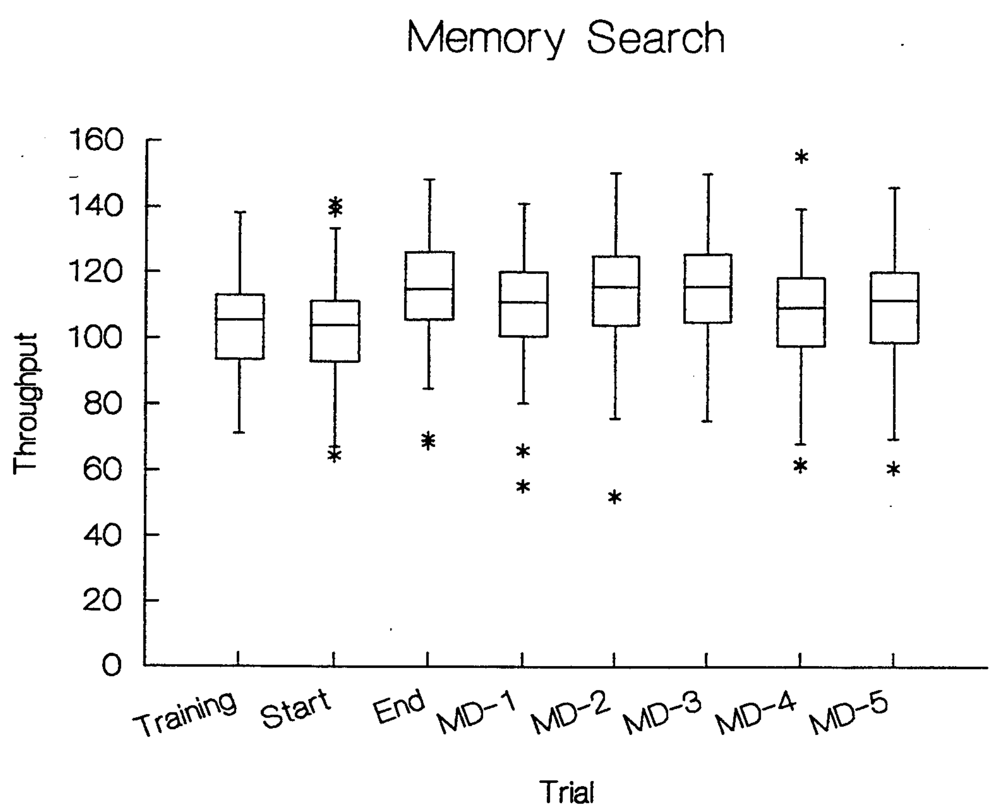

Continuous Recognition
100
90
80
70
60
Percent Correct
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial



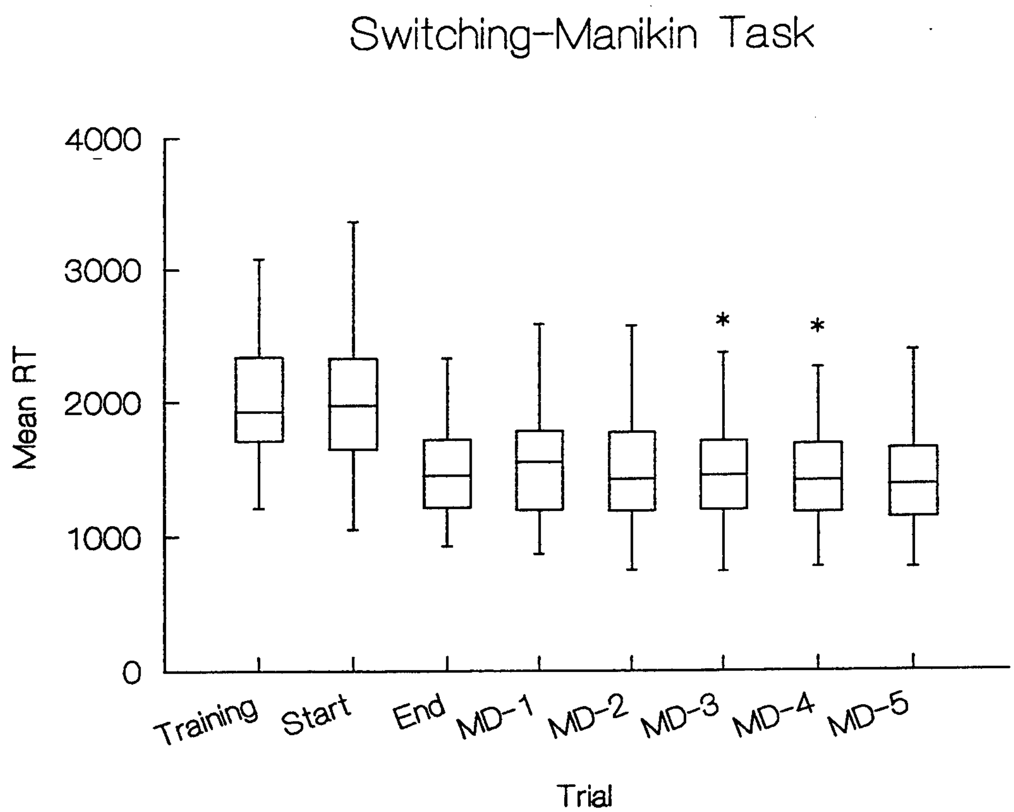
Continuous Recognition
Throughput
120
100
80
60
40
20
0
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

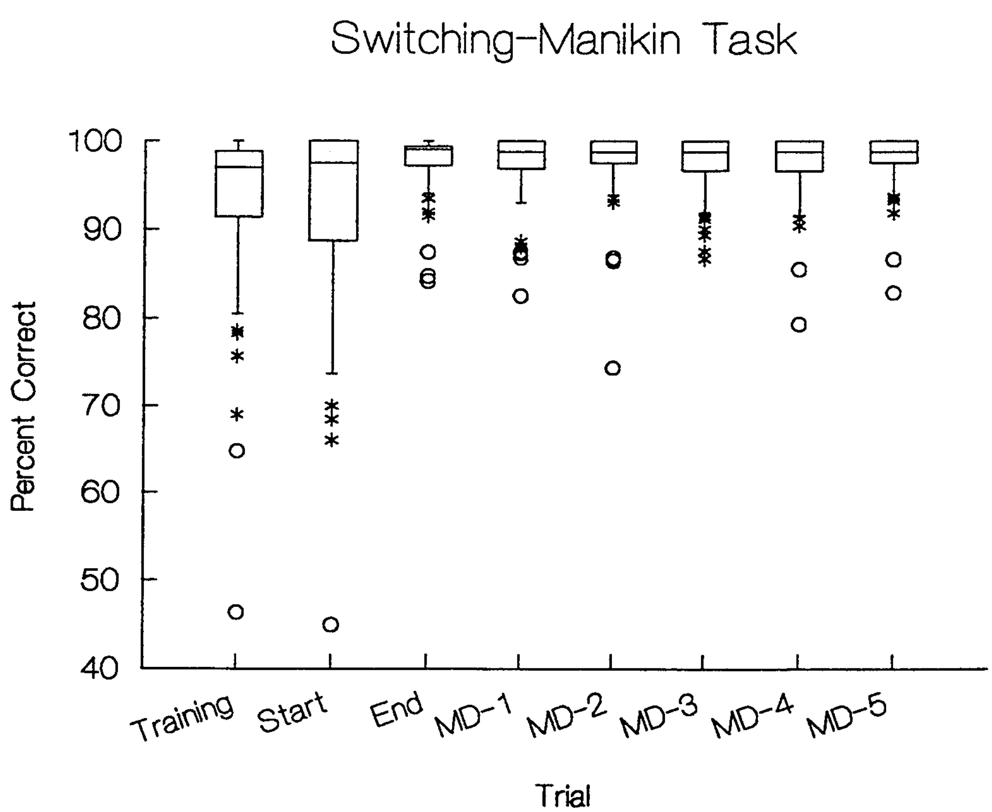

Memory Search
1200
1000
800
600
400
200
0
Mean RT
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial


Memory Search
100
95
90
85
80
Percent Correct
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

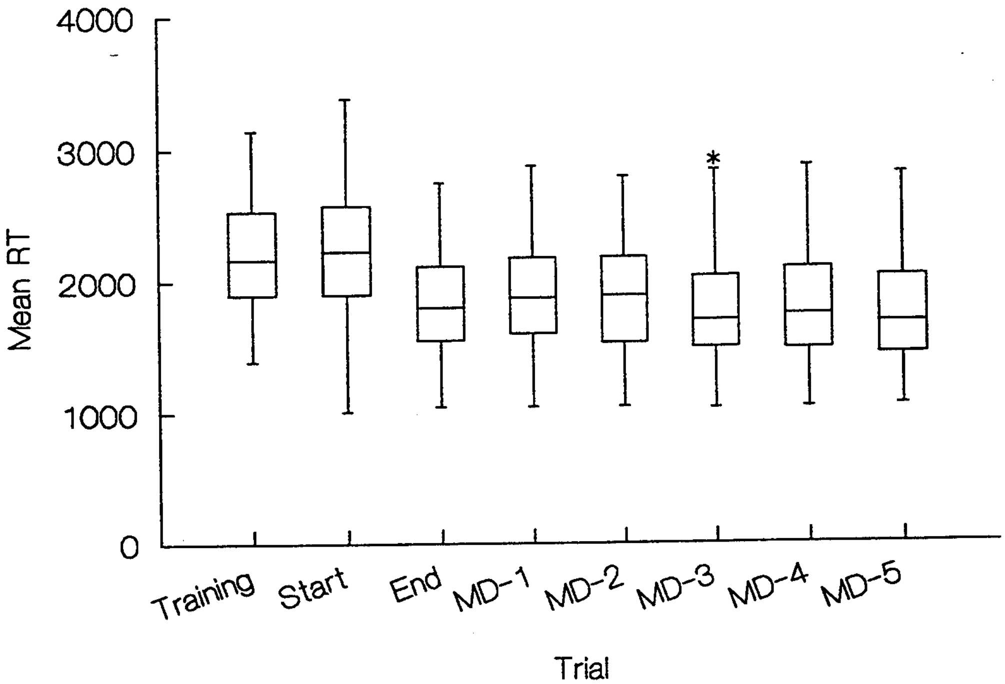


Memory Search
Throughput
160
140
120
100
80
60
40
20
0
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

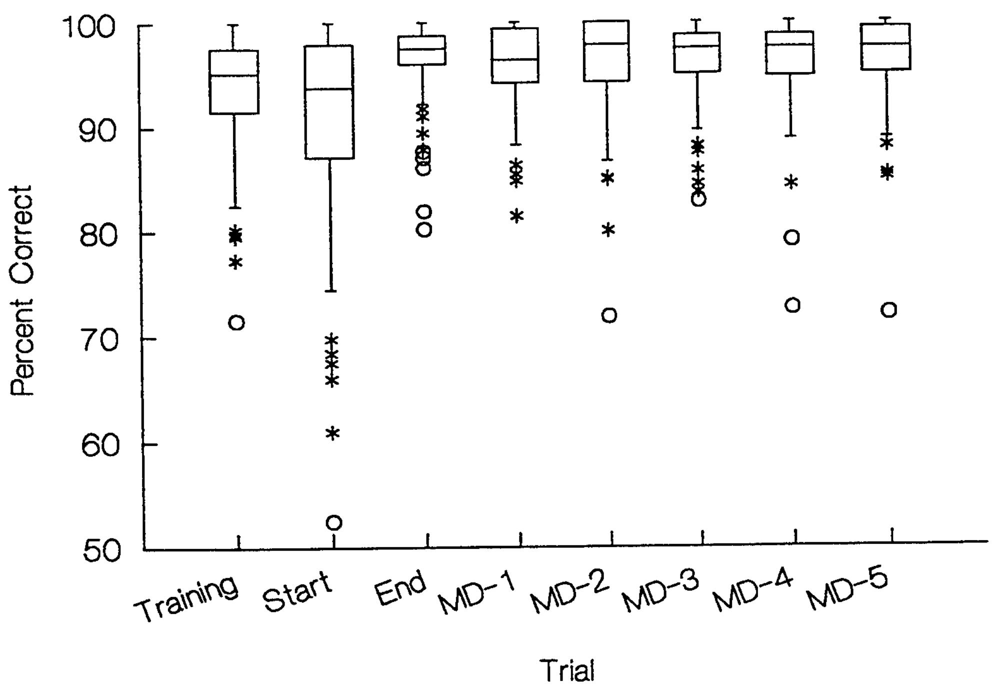

Switching-Manikin Task
Mean RT
4000
3000
2000
1000
0
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

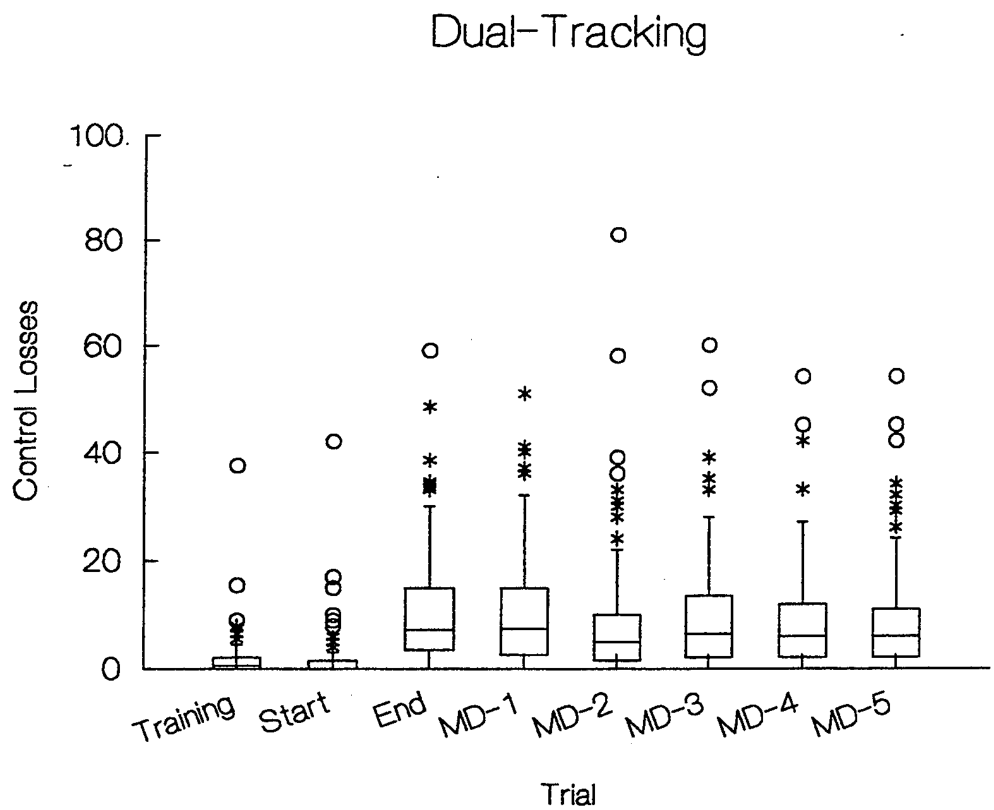

Switching-Manikin Task
Percent Correct
100
90
80
70
60
50
40
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

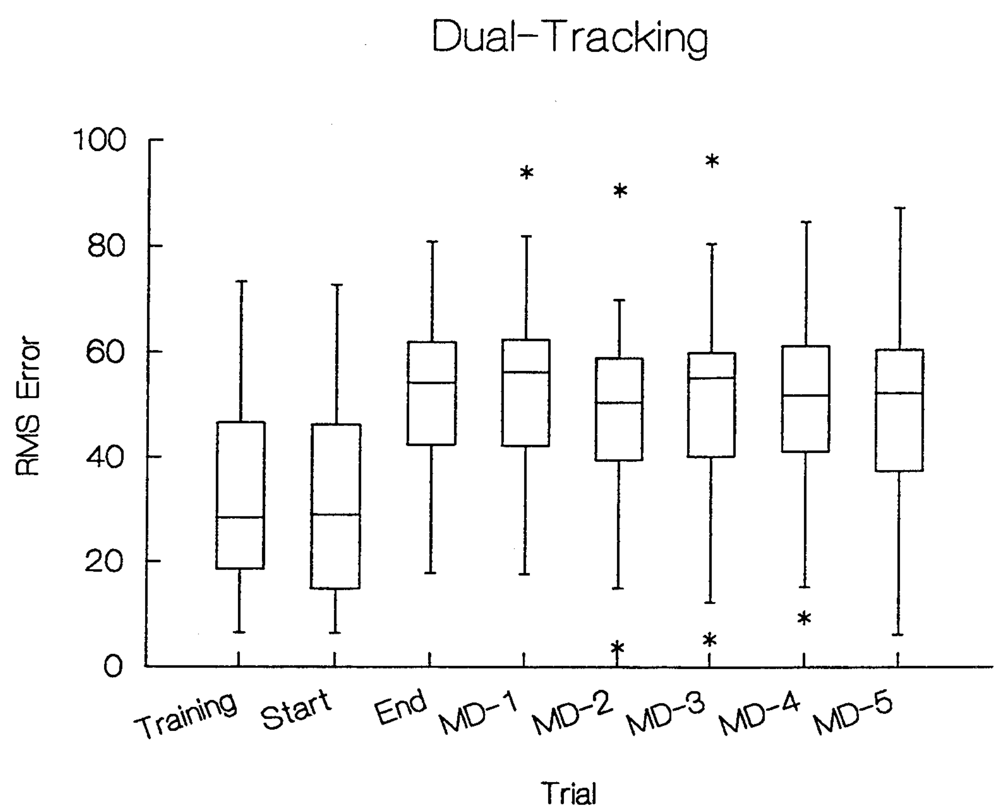
Switching-Mathematical Processing

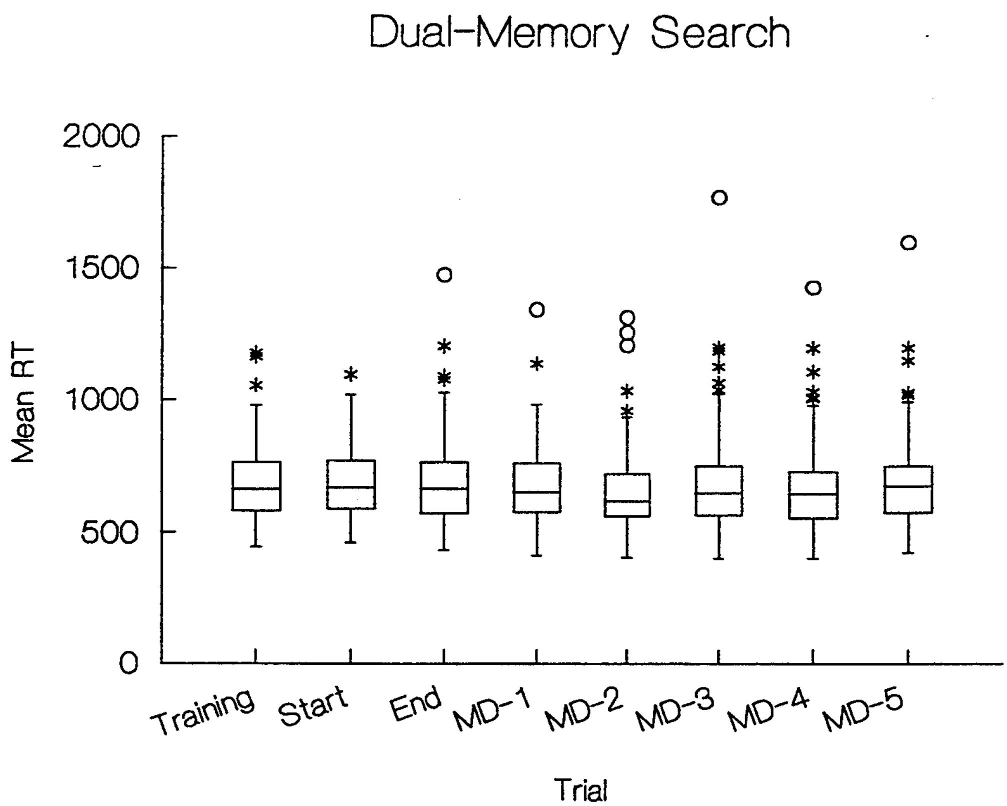


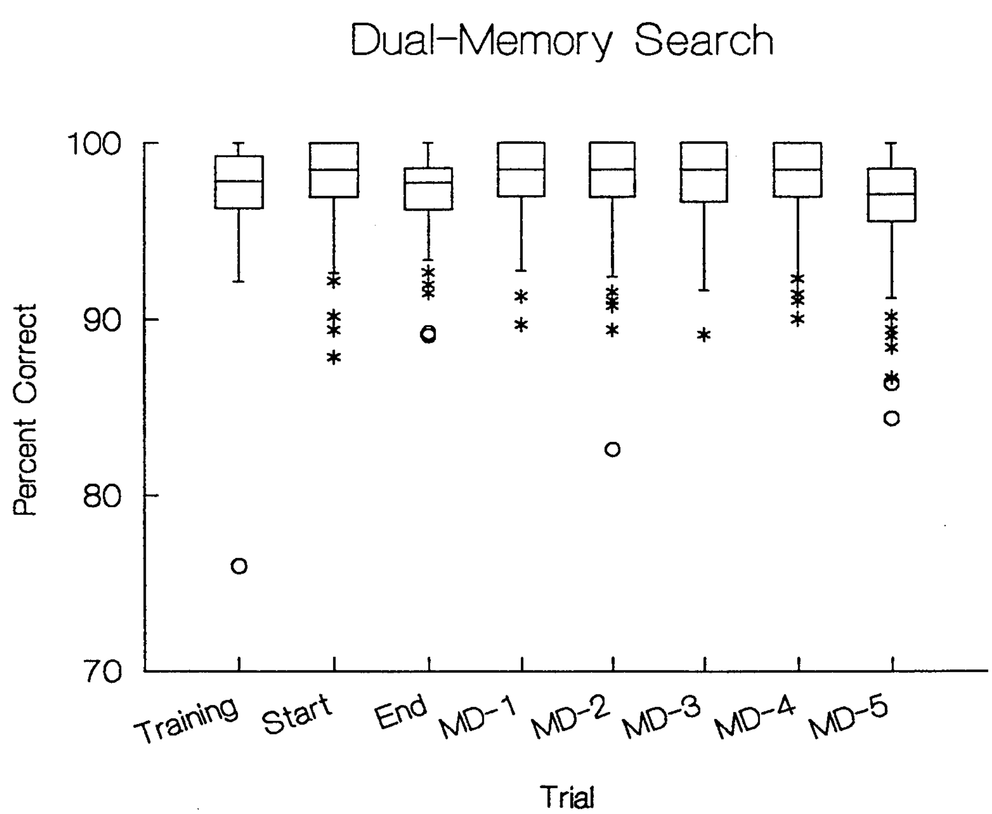
Switching-Mathematical Processing

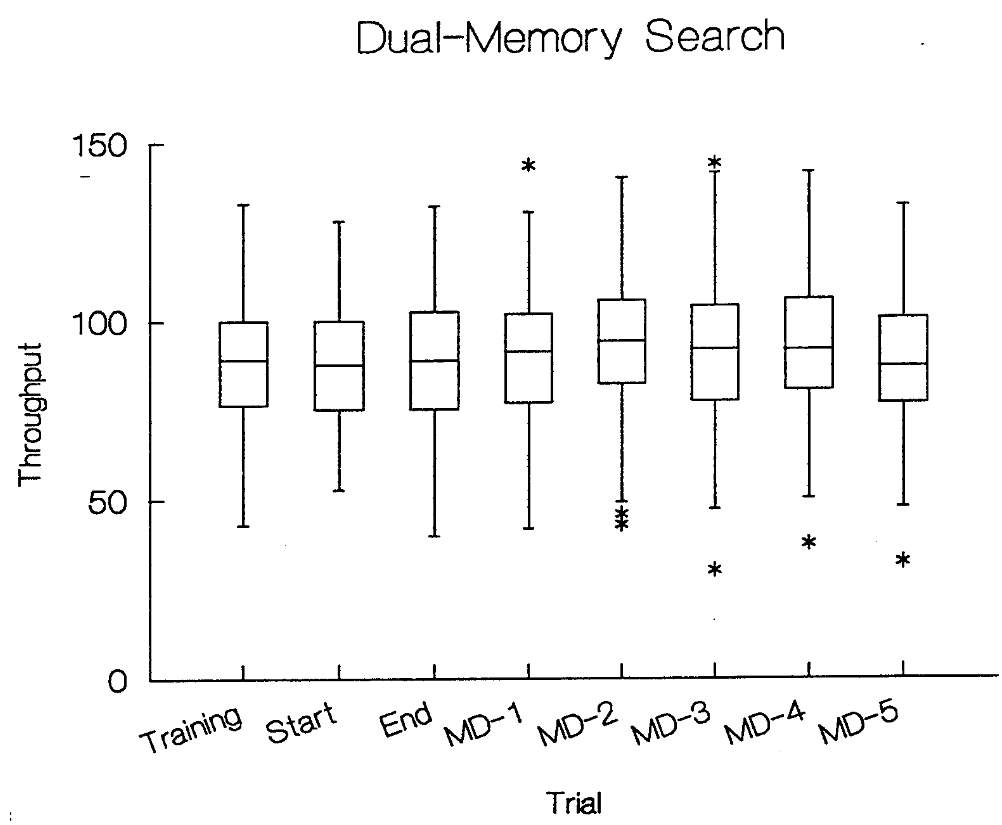


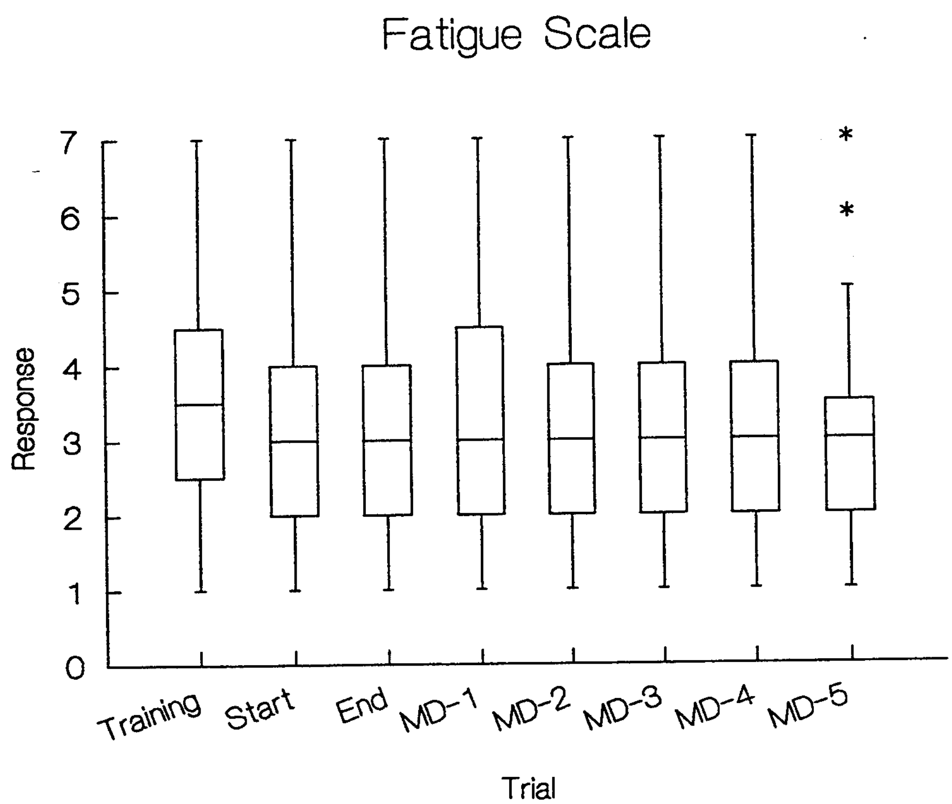

Dual-Tracking
Control Losses
100.
80
60
40
20
0
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

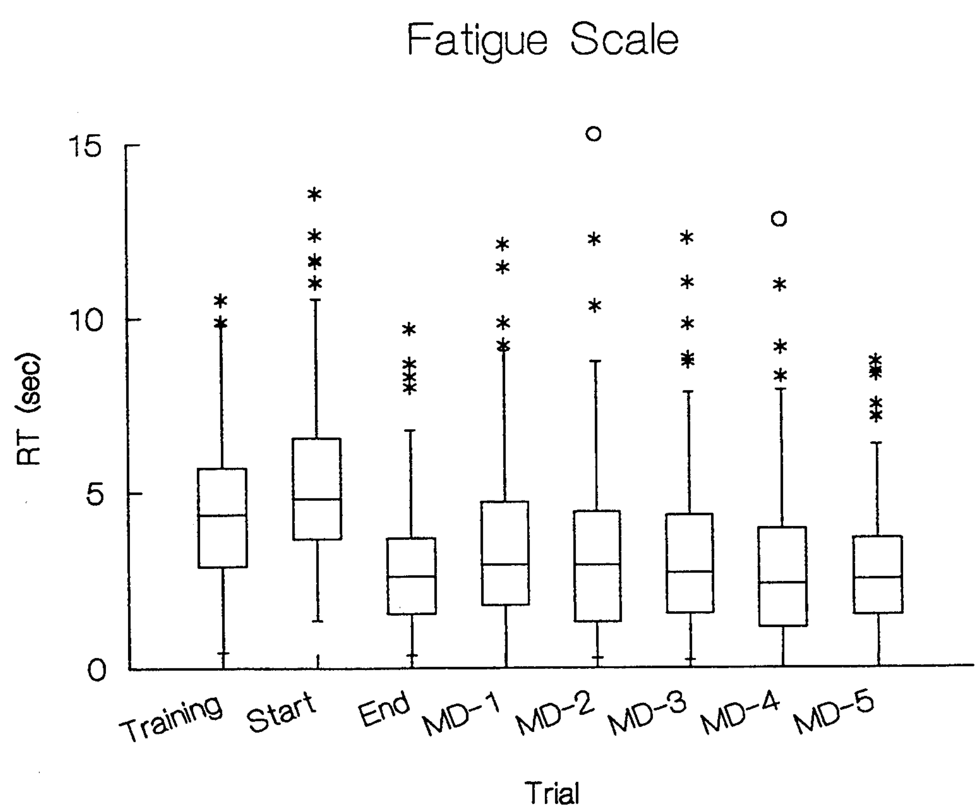

Dual-Tracking
RMS Error
100
80
60
40
20
0
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial


Dual-Memory Search
Mean RT
2000
1500
1000
500
0
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial


Dual-Memory Search
Percent Correct
100
90
80
70
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial


Dual-Memory Search
150
100
50
0
Throughput
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial


Fatigue Scale
Response
0
1
2
3
4
5
6
7
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial


Fatigue Scale
RT (sec)
0
5
10
15
Training
Start
End
MD-1
MD-2
MD-3
MD-4
MD-5
Trial

APPENDIX J

SUBJECT DEBRIEFING

SUBJECT DEBRIEFING QUESTIONNAIRE - SUBJECT NO. ____

This questionnaire is designed to find out about your experience as a subject in this experiment. It will take less than 10 minutes to fill out. Please be very honest in all your responses and don't be afraid of offending the research team. Your thoughts and comments are needed and appreciated.

1. What is your overall impression of being a subject in this experiment?

   Would you do it again or recommend it to a friend?

2. What was(were) your most preferred task(s)? Why?

   Least preferred? Why?

3. Did you receive enough instructions about the tasks?

   Did you ever depart from a task's standard instructions (e.g., used wrong fingers on the keys) or forget instructions?

4. Did you thoughtfully and accurately respond on the MOOD and FATIGUE scales?

   Do you have any other comments about the MOOD and FATIGUE scales?

5. Please note any special strategies (or "tricks") you developed to improve your performance on the tasks:

a. Tracking

b. Matrix

c. Sternberg Memory

d. Continuous Recognition

e. Switching

f. Dual (Tracking and Memory)

6. Do you have any comments about the testing environment (e.g., lighting, disturbances from other subjects, etc.)?

7. How were you treated by the experimental staff? Please be honest.

SUBJECT DEBRIEFING SURVEY

(NUMBER OF RESPONDENTS = 79)

1. What is your overall impression of being a subject in this experiment?

Positive - 71 (interesting, challenging, enjoyable, good, easy, great, fine)
Negative - 14 (boring, monotonous, tiresome, too long)
Neutral - 3
Did it for the money - 6

Would you do it again or recommend it to a friend?

Yes - 68
Probably - 1
Yes, with conditions - 10

2. What was (were) your most preferred task(s)? Why?

Task	Number	Why?	Number
TRK	13	challenging	8
		fun	3
		not sure why	2
MTX	10	challenging	3
		fun	4
		easy	1
		not sure why	1
STN	20	easy	15
		challenging	1
CRC	18	easy	7
		like numbers	2
		challenging	3
		fun	2
		good at it	1
SW	18	challenging	9
MAN	11	easy	3
MTH	8	fun	4
		quick	1
		interesting	3
		entertaining	2
		good at it	1
		improved most	1
DUL	16	challenging	9
		easy	3
		fun	3

2. (continued)
Least preferred? Why?

Task	Number	Why?	Number
TRK	28	too hard	15
		pain	1
		frustrating	5
		long	2
		boring	1
MTX	13	too hard	7
		boring	2
		software	1
STN	10	too easy	5
		boring	4
		too long	2
		silly errors	1
CRC	10	too hard	6
		boring	1
SW	12	too hard	5
MAN	3	probs switching	3
MTH	1	hate math	1
		too easy	1
		boring	1
		buttons reversed	1
		not sure	1
DUL	16	too hard	8
		pain	2
		too long	2
		boring	2
		software mistakes	1

3. Did you receive enough instruction about the tasks?

Yes - 75
No - 4 (Switching - 2; RT vs. accuracy tradeoff - 1; Dual when it got more difficult - 1)
Mostly - 1

Did you ever depart from a task's standard instructions (e.g., used wrong fingers on the keys) or forget instructions?

Yes - 32 (MAN bkwds - 3; MTX bkwds - 3; thought "I" = identical - 1; MTH bkwds - 3; used wrong hand on MOOD - 1, SW - 1; used wrong fingers on SW first time - 2; used wrong fingers on DUL - 1; did not use fingers on STN - 1)
Once - 9
Twice - 4
Finger slipped - 2
Yes, during first week - 11
No - 30
Not to my knowledge - 4
After the time off - 1

4. Did you thoughtfully and accurately respond on the MOOD and FATIGUE scales?

Yes - 50
Usually - 19
Half of the time - 1
Yes on FATIGUE, not always on MOOD - 1
Sometimes hit wrong key on MOOD - 13

Do you have any other comments about the MOOD and FATIGUE scales?

Did not cover all possibilities - 10
Too easy to memorize - 3
Different words for same mood - 4
Fatigue level changed during session - 6
Not descriptive - 3
Mood scale was not too valid - 2
Should ask MOOD also after session - 2
Always responded the same on MOOD - 2
Covered a wide range - 2
(1 each) - Ambiguous words
- Helped determine own mood
- Need last one separate from "ready to drop"
- Interpretation of words changed during experiment
- Fatigue scale was too descriptive
- Should do it once for both tests (on two-a-day sessions)
- There should be one more adjective between options 2 and 3 in Fatigue scale
- Mood scale was not meaningful
- Mood scale was long and tedious
- Scales did not account for gender differences
- Negative adjectives did not reflect actual mood
- Got tired of responding
- Hard matching responses to actual mood
- Became a routine
- Mood scale was a little vague

5. Please note any special strategies (or "tricks") you developed to improve your performance on the tasks:

a. Tracking
None - 33
Quick, jiggling motions - 6
Small movements - 5
Relax - 5
Move trackball oppositely as soon as cursor changed direction - 5
Bounce around center - 3
Rest arm on chair arm - 2
Try to keep cursor to one side of center - 2
Use two fingers - 2
Move trackball slowly - 2
Avoid jerking motions - 2
(1 each) - Keep hand steady
- "Cool out"
- Sharp, jerking motions
- "Moving then stopping ball fast"
- Develop a sense of direction, then move
- Blinked less when wearing glasses instead of contacts

5a. (continued)

- "Feathering or brushing"
- "Get pissed at it"
- Do not try to stay in middle
- Try to anticipate amount of roll needed to handle cursor
- Sit back and use one finger

b. Matrix

None - 19
Look at overall picture - 18
Look at 2 or 3 boxes together - 14
Go fast/use intuition - 11
Look at corners - 7
Look for patterns or designs (e.g., "L shapes") - 9
Look at edges - 3
(1 each)
- Imagined picture
- Dark and light patterns
- Memorize row and column numbers
- Looked for squares connected diagonally
- Looked for partial matches <u>and</u> go fast
- Filled in spaces as if boxes fell when rotated(?)
- Looked at blanks
- Look at from periphery
- Anticipate 4 positive responses in a row
- Respond "same" unless I was sure 100% it was different
- Looked for two blocks I could identify with
- Slow down

c. Sternberg

None - 22
Repeat positive set - 28
Created word from positive set - 22
Keep finger(s) slightly above keys - 2
Try not to anticipate - 2
Make rhyme out of letters - 2
(1 each)
- Memorize in groups of two
- Concentrate on fingers
- Do not make word from positive set
- Visualize letters when not on screen
- Time eye blinks
- Try to stay distracted
- Try not to look at positive set long
- Concentrate and slow down
- Create a phrase from letters using positive letters as word initials
- Take time to arrange positive set mentally in a comfortable way
- Related letters to words or names
- Go fast

d. Continuous Recognition

None - 27
Say numbers aloud or in mind - 15
Concentrate on bottom number - 16
Get into a rhythm - 5
Go fast - 7
Look at whole fraction - 3
Thought of numbers as scrolling - 3
(1 each)
- Intuition
- Rotated head, made numbers go left to right, not top to bottom

5d. (continued)

- Saw fraction as 3 numbers scrolling upward
- Repeating in mind "yes" or "no"
- Memorize bottom number and type of response (correct or incorrect)
- Concentrate on fraction line
- Use vision, don't memorize numbers
- Focus on bottom number and go fast
- Slow down
- Say "yes" and "no" to myself

e. Switching

None - 39

Go fast/use intuition - 2

Look at bar at bottom - 2

(1 each)
- Maintain concentration
- "Flow your eyes"
- See bottom bar in periphery

Manikin:
- Imagined/moved body in position - 9
- Memorized four (?) possibilities - 8
- Matched response keys to Manikin's orientation - 2
- Used opposites - 3
- Memorized specific possibility - 2
- Checked always manikin's right hand for a match. If no match was found, response was "left"

Math Processing:
- Used ballpark estimate - 5
- Look for all "+" or all "-" signs - 4
- Look for "+ 4" or greater at right - 2
- (1 each)
  - Add right to left
  - Do subtraction first
  - Watch for repeats
  - Use common sense
  - Approximate addition result
  - Go fast

f. Dual (Tracking and Memory)

None - 31

Stress TRK; do STN in periphery - 20

Same as a above - 2

Same as c above - 4

Same as a and c above - 8

Stress STN; do TRK in periphery - 7

(1 each)
- Do both simultaneously
- Concentrate on center of screen
- Move cursor slowly/keep it in middle
- Slow down
- Stress TRK and create a phrase from letters
- Concentrate on TRK the second half of test

6. Do you have any comments about the testing environment (e.g., lighting, disturbances from other subjects, etc.)?

No comment - 22
Okay - 18
Problem with screen glare - 13
Preferred station #1 - 2
Arm hurt - 2
Isolate subjects more - 2
Cold room - 1
Back hurt - 1
Did not like having to use right hand - 1
Uncomfortable trackball - 1
Adjust screen brightness - 1
Did not like being watched behind my back - 1
Perfect - 1
Different trackball sensitivity - 1
Trackball was unstable - 1
Stuffy room -1
Distracted by subjects coming and going- 8
" " " hitting keys - 2
" " people talking - 1
" " computer beep - 3
" " sound - 3
" " telephone -1
" " noise from next room - 1
" " experimenters - 2
" " dripping faucet - 3
" " elevator - 1
" when moving things above my head - 1

7. How were you treated by the experimental staff? Please be honest.

Positive - 76
Helped with schedule - 9

Negative - 3