File Cy - Do Not Remove

November 1983
Final Report

DOT HS-806-459

U.S. Department of Transportation
National Highway Traffic Safety Administration

Drunk Driving Warning System (DDWS)

Volume I
System Concept and Description

R. Wade Allen
Anthony C. Stein
Leland G. Summers
Marcia L. Cook

Systems Technology, Inc.
13766 S. Hawthorne Boulevard
Hawthorne, California 90250

Contract No. DOT-HS-8-02052

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS-806 459		

4. Title and Subtitle	5. Report Date
DRUNK DRIVING WARNING SYSTEM (DDWS) Volume I: System Concept and Description	November 1983
	6. Performing Organization Code
7. Author(s) R. Wade Allen, Anthony C. Stein, Leland G. Summers, and Marcia L. Cook	8. Performing Organization Report No. STI TR-1136-1-I
9. Performing Organization Name and Address Systems Technology, Inc. 13766 S. Hawthorne Boulevard Hawthorne, California 90250	10. Work Unit No. (TRAIS) 11. Contract or Grant No. DOT-HS-8-02052
12. Sponsoring Agency Name and Address Department of Transportation National Highway Traffic Safety Administration Washington, D. C. 20590	13. Type of Report and Period Covered Final Report October 1978 - November 1983 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The Drunk Driving Warning System (DDWS) is a vehicle-mounted device for testing driver impairment and activating alarms. The driver must pass a steering competency test in order to drive the car in a normal manner. The emergency flasher system operates when the test is failed; and if the car is driven faster than 10 mph, the horn honks periodically, once per second. The DDWS concept and hardware were developed under earlier contracts. The purpose of this work was to evaluate the feasibility of the concept, both in its sensitivity to alcohol impairment and in terms of various practical considerations in assigning the device to DWIs as a judicial sanction in a probationary setting.

Volume I of the final report focuses on optimizing and validating the sensitivity of DDWS to alcohol impairment. Background on DDWS development and mechanization is given, followed by reanalysis of past performance data and optimization of test parameters. The final system configuration was tested in a laboratory experiment. The discriminability of the test strategy to BAC (blood alcohol concentration) is confirmed. The ability of DDWS to discriminate impaired performance in a driving simulation is also demonstrated.

Based on the analysis, optimization, and experimental validation of DDWS performance, recommendations are given for test application and training procedures. These recommendations apply to the field test evaluation described in Volume II of this report.

17. Key Words	18. Distribution Statement
Drunk driving, vehicle instrumentation, tracking performance, psychomotor testing.	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified		

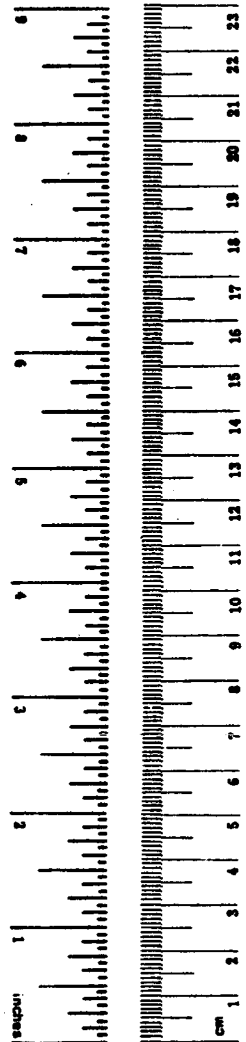
Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

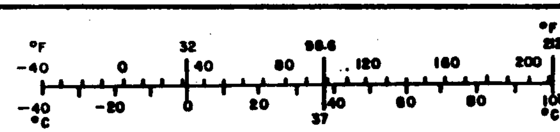
Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



FOREWORD

A number of individuals contributed to the work described in this volume and its preparation. Drs. Thomas G. Ryan and Marvin M. Levy, who served as NHTSA Contract Technical Managers at various stages of the project, made important editorial contributions. The Honorable Roy Carstairs and Sherman Smith, judges with the West Los Angeles Municipal Court assigned convicted drunk drivers (DWI's) to serve in the validation experiment as a condition of probation.

Several individuals at Systems Technology, Inc. also provided invaluable support. Henry R. Jex, one of the original developers of the CTT (Critical Task Tester), and Richard A. Peters, developer of the DDWS (Drunk Driving Warning System) apparatus gave significant technical assistance. Zareh Parseghian, Robin L. Karl, and James Nagy supported the preparation of experimental apparatus and conduct of the validation experiment. Raymond E. Magdaleno prepared data analysis software and conducted much of the CTT data reanalysis. Final preparation of this document was admirably handled by the STI publications staff including Winifred A. Reaber, Linda L. Huffman, Charles W. Reaber, and Jon Petitjean.

ADDENDUM

As part of the National Highway Traffic Safety Administration's comprehensive approach for reducing the drunk-driving problem, information is being collected and analyzed on the potential utility of in-vehicle systems designed to deter or prevent driving while intoxicated. In this study, an innovative sanction, called a Drunk Driving Warning System (DDWS) was pilot tested in a judicial setting in California. Conceptually, with a DDWS, a driver takes an in-vehicle test for the presence of alcohol. If the test is passed, the car operates normally. However, when the test is taken and failed or not taken at all, the parking lights flash and, if the car is driven over a set speed (e.g., 10 mph), the horn honks, warning others, including the police that the driver is impaired.

An objective of this study was to improve the adequacy of the performance test component-- called the Critical Tracking Task (CTT)--that was incorporated as part of the DDWS examined in this project. As reported in this volume, the contractor developed a training, testing, and scoring procedure, for use with drivers categorized as heavy drinkers, that couples a low sober-test failure rate (two-three percent failures at 0.0% Blood Alcohol Concentration (BAC)) with a high test-failure rate when intoxicated (approximately 80 percent test failures at about 0.15% BAC). However, at 0.10% BAC--the presumptive level in the majority of states--this procedure results in driver failures at about a 33 percent rate. Other procedures, to increase the test-failure rate with the CTT at 0.10% BAC, are currently being investigated in-house by NHTSA.

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
	A. Background and Overview	1
	B. Objectives of DDWS Project	2
	C. Objectives of Volume I Report	2
	D. Volume I Report Outline	3
II.	BACKGROUND	5
	A. Overview	5
	B. Early NHTSA Research	5
	C. Critical Tracking Task (CTT)	7
	D. Drunk Driving Warning System	8
III.	CRITICAL TRACKING TASK (CTT) DATA ANALYSIS	9
	A. Overview	9
	B. Average CTT Performance	9
	C. Variability in CTT Performance	12
	D. Summary and Conclusions	14
IV.	CTT TEST OPTIMIZATION	15
	A. Overview	15
	B. Test Strategy	15
	C. Pass Criteria	18
	D. Training	21
	E. Summary and Conclusions	22
V.	CTT LABORATORY VALIDATION EXPERIMENT	23
	A. Experimental Design	23
	B. Procedures	25
	C. Results	27
	D. Summary, Conclusions, and Recommendations	30
	E. Field Test Recommendations	32

TABLE OF CONTENTS (Continued)

		Page
VI.	DRUNK DRIVING WARNING SYSTEM (DDWS) SYSTEM CONFIGURATION	33
	A. Overall Installation	33
	B. Basic Operation	36
	C. Anticircumvention	38
VII.	EPILOGUE	40
REFERENCES		41
APPENDIX A.	CTT DATA ANALYSIS	A-1
APPENDIX B.	TEST DISCRIMINABILITY ANALYSIS	B-1
APPENDIX C.	MODEL VALIDATION AND ANALYSIS	C-1
APPENDIX D.	VALIDATION EXPERIMENT DETAILS	D-1
APPENDIX E.	CRITICAL TRACKING TASK TRAINING AND PASS LEVEL SELECTION	E-1
APPENDIX F.	DDWS SYSTEM CONFIGURATION	F-1

LIST OF FIGURES

		Page
1	Effect of BAC on Normalized CTT Score....................	11
2	Effects of Alcohol Intoxication on DDWS Test Failure for one out of n Test Strategies............	17
3	CTT Learning Curve Averaged Across All Subjects..........	28
4	CTT Alcohol Impairment Discriminability Validation Data Compared with Statistical Model Prediction..........	29
5	Comparison of Driving Simulator Performance CTT Failure Rate..	29
6	Comparison of CTT and Simulator Performance with Real World Accident Rates (Real World Data Adopted from Hurst, 1973 and Voas, 1974)........................	31
7	CTT Discriminability of Impaired Driving Simulator Performance..	31
8	Dash Mounted CTT Meter..................................	34
9	Trunk Mounted Electronics Module and Data Logger.........	35
10	The Seat Weight Sensor System (Underside View of Driver Seat)..	37
A-1	Critical Task Score as a Function of BAC Across Several Experiments. Data Normalized to Account for Differences in Task Mechanization, Procedures, etc...	A-4
A-2	Cumulative Distribution Functions of Normalized λ_c Scores from the Dunlap Experiments......................	A-7
A-3	BAC Effect on the Differential Mean and Variability of CTT Scores..	A-12
B-1	CTT Gaussian Score Distributions According to Appendix A Model...	B-3
B-2	Discriminability of Average of n Trials Strategy Applied to Dunlap Data Base for α = 0.025................	B-7
B-3	Discriminability of One Out of n Trials (1/n) Strategy Applied to Dunlap Data Base for α = 0.025.......	B-9
B-4	Discriminability of m out of n Trials Strategy Applied to Dunlap Data Base for α = 0.025................	B-13
B-5	Discriminability of a Simple Sequential Sampling Strategy Applied to the Dunlap Data Base for α = 0.025...	B-17
B-6	Discriminability of a Likelihood Ratio Strategy Applied to the Dunlap Data Base for α = 0.025...........	B-21
B-7	Comparison of Test Strategy BAC Discriminability.........	B-23

LIST OF FIGURES (Concluded)

		Page
C-1	Model Analysis of the Effects of Number of Trials (n) and Sober Failure Rate on the BAC Discriminability of the One of n (1/n) Test Strategy; Model Assumes Individualized Pass Levels	C-2
C-2	The Effect of Setting Different Levels of α on Individualized Pass Criteria Levels	C-4
C-3	Discriminability Differences Between Individualized and Universal Pass Criteria Levels	C-7
C-4	Hypothetical CTT Score Distributions Based on Appendix A Model Illustrating the Consequence of Universal Pass Level Criteria	C-8
D-1	Interactive Driving Simulator	D-3
D-2	Typical Subject's Participation	D-5
D-3	Typical Experimental Day	D-6
E-1	Average Learning Trends for Dunlap CTT Data	E-3
E-2	Comparison of the Second Highest Training Block Mean Score (Last Training Session) with Mean Experimental Performance	E-5
E-3	Pass Criterion Analysis	E-8
E-4	Learning Curve for a Subject Who Received Delayed Monetary Rewards for Task Performance in Addition to Immediate Wait Time Penalties	E-11
E-5	Learning Curve for a Subject Receiving No Monetary Rewards for Task Performance	E-12
E-6	Learning Curve for a Second Subject Receiving No Monetary Rewards for Task Performance	E-13
F-1	Critical Task Interlock - Assemblies and Interfaces	F-2
F-2	Front Panel of Critical Task Interlock and Data Logger	F-5
F-3	Critical Task Interlock Display and Steering Wheel Pickoff	F-7

LIST OF TABLES

		Page
1	Component Variability Summary for CTT Data. Standard Deviation Expressed as a Percentage of Sober Mean Score. Data Obtained Through Analysis of Variance Procedures.....	13
2	Experimental Design....................................	24
A-1	Analysis of Variance for CTT Data........................	A-5
A-2	Subject-by-Subject Regression Analysis of Dunlap Data.....	A-10
B-1	Average of n Trials Test Strategy Applied to Dunlap Data Base for α = 0.025................................	B-6
B-2	1 Out of n Trials Test Strategy (1/n) Applied to Dunlap Data Base for α = 0.025............................	B-8
B-3	m Out of n Trials Test Strategy Applied to the Dunlap Data Base for α = 0.025............................	B-12
B-4	Simple Sequential Test Strategy Applied to the Dunlap Data Base for α = 0.025............................	B-16
B-5	Likelihood Ratio Sequential Test Strategy Applied to the Dunlap Data Base for α = 0.025....................	B-20
C-1	Discriminability Analysis of Phase I Dunlap Data (Ref. 6) for Universal vs. Individualized Pass Levels Using Regression Model Parameters from Appendix A.........	C-5
E-1	Trials and Rewards.....................................	C-2
F-1	DDWS Recorded and Computed Data..........................	F-9

SECTION I

INTRODUCTION

A. BACKGROUND AND OVERVIEW

National Highway Traffic Safety Administration (NHTSA) research and development personnel have been investigating the potential of several in-vehicle devices designed to deter persons who are alcohol impaired from driving. These devices require that the driver pass a behavioral test before operating the vehicle. A behavioral test is defined here as any manual dexterity task involving reaction time, divided attention, short-term memory, eye-hand coordination, etc. NHTSA researchers have also been investigating potential legal and public acceptance issues associated with employment of such a device on vehicles operating on Federal, State and local highway systems. These in-vehicle devices are presently envisioned for use only among persons convicted of Driving While Intoxicated (DWI).

Currently, NHTSA is investigating the potential of an in-vehicle Drunk Driving Warning System (DDWS) concept, which permits the vehicle to be operated under conditions when the behavioral test required by the system is not passed or is not attempted. However, when the vehicle is operated under such conditions, other drivers (and police) are alerted by blinking emergency lights and at speeds of 10 miles per hour (mph) and above, by the horn honking intermittently. NHTSA has abandoned the concept of an interlock device which makes starting/operating a vehicle "impossible" when the performance level required by the device is not achieved. Practical reasons for abandoning the interlock concept included such requirements as emergency operation of the vehicle (e.g., neighbor rushing a pregnant woman to the hospital) or quickly removing the vehicle from the middle of a busy freeway or intersection if its engine has stalled.

This report describes preliminary work on a pilot field test of the DDWS concept under Contract DOT-HS-8-02052, "Field Test of the Drunk Driving Warning System." This feasibility test project was designed to investigate the utility of the DDWS concept in a judicial setting in which a person convicted of a second or subsequent DWI is afforded the alternative of driving a vehicle equipped with the DDWS, rather than undergoing the more common sanctions, e.g., license suspension, jail, etc. The project was designed to provide credible information on the reliability of the DDWS, the procedure for its implementation in a judicial setting, and the ability of the DDWS to deter drunk driving trips. Data from the project will be used by NHTSA in deciding whether the DDWS concept should be implemented and tested on a wider scale.

DDWS project background and results are described in a two volume final report. This report, Volume I, reviews research leading to the DDWS concept, and describes the current DDWS being tested among convicted drunk-drivers in California.

B. OBJECTIVES OF DDWS PROJECT

The overall objectives of the DDWS project were to: (1) complete developmental testing of the DDWS, especially its test component Critical Tracking Task (CTT), (2) prepare the DDWS for field testing in a judicial setting with convicted drunk drivers, and (3) conduct a field test of the system to investigate its practicality, public acceptability and effectiveness when used with convicted drunk drivers.

C. OBJECTIVES OF VOLUME I REPORT

This Volume I of the final report is devoted to the first two project objectives above. More specifically, the objectives of this Volume I report are as follows: (1) review the extensive amount of prior research available on in-vehicle alcohol safety devices, especially the Critical Tracking Task (CTT); (2) identify the alcohol sensitivity properties of the CTT and derive a statistical model of CTT performance that will allow analysis and optimization of CTT discriminability and ultimately DDWS ability to deter drunk driving trips; and

(3) report on laboratory and driver simulator tests designed to validate pass/fail criteria and test strategies (number of test trials permitted before a test failure is determined).

D. VOLUME I REPORT OUTLINE

Section II reviews background research on in-vehicle alcohol safety devices leading up to the selection of the Critical Tracking Task (CTT) as the test component of the Drunk Driving Warning System (DDWS). Early in-vehicle devices were conceived as ignition interlocks but practical considerations (i.e., emergency operation and safety) eventually led to a warning system concept.

Section III reviews earlier research which established the alcohol sensitivity and general statistical properties of the CTT. These analyses showed that CTT performance was degraded approximately 10 percent at a Blood Alcohol Concentration (BAC) level of 0.10. Secondly, they showed a performance variability by subjects from one test trial to the next of about the same magnitude as the performance decrement at 0.10 BAC, which dictated the use of multiple trial test strategies to optimize CTT discriminability. Thirdly, they showed that sober performance variability between subjects was also about the same magnitude as the within-subject (trial-to-trial variability, which dictated setting individualized pass-fail scores to further optimize CTT discriminability.

Section IV reviews the optimization of CTT test parameters and training procedures. Optimization of test parameters included development of a multiple test strategy (i.e., several attempts permitted in order to pass the test), and development of individual pass/fail scores. Test parameter optimization was judged from the viewpoint of minimizing the chances of sober drivers failing the CTT test and drunk drivers passing the test. Training procedures were developed to reliably establish each subject's sober performance capability on the CTT, which would then be used to set his pass-fail score during formal tests on the device.

Section V summarizes laboratory and driver simulator experiments designed to validate optimized CTT test strategy parameters and training procedures. Results of these experiments demonstrated that CTT discriminability was highly reliable. Analysis of corresponding driver simulator data showed a high correlation with CTT performance data under sober and alcohol impaired conditions.

Section VI reviews the current DDWS system configuration, including its physical installation in a vehicle and its basic operating characteristics.

Section VII summarizes the purpose of this Volume I. It also outlines the purpose and objectives of the Volume II report (i.e., report results from field test of the DDWS concept in a judicial setting among convicted drunk drivers).

More detailed technical background is provided in several appendices to this volume. Appendices A-C analyze the statistical properties of CTT scores and optimize CTT discriminability. Methods for the validation experiment are summarized in Appendix D. Details on training procedures are given in Appendix E. Finally, a detailed description of the DDWS apparatus is given in Appendix F.

SECTION II

BACKGROUND

A. OVERVIEW

An in-vehicle alcohol safety device, in order to be a successful deterrent to drunk-driving, must meet two criteria: (1) provide a sensitive measure of alcohol impairment, and (2) be resistant to circumvention. NHTSA early research on such a device is described in Article B. Article C provides an overview of the Critical Tracking Task (CTT) eventually selected for incorporation into a vehicle-mounted Drunk Driving Warning System (DDWS). An overview of the DDWS is provided in Article D.

B. EARLY NHTSA RESEARCH

NHTSA research on in-vehicle alcohol safety devices began in 1970. Early research studies were conducted for NHTSA by the Department of Transportation's Transportation Systems Center (TSC). Eight alcohol safety devices were tested by TSC for their ability to discriminate between sober and alcohol impaired subjects, and for their ability to withstand circumvention. Each of the devices tested required the user to pass a behavioral test involving reaction time, divided attention, short-term memory and/or eye-hand coordination. After initial screening, four of these devices were selected for comprehensive testing (i.e., Quickey, Phystester, Reaction Analyzer, Complex Reaction Tester). The results of these latter tests were not particularly promising. Failure rates at Blood Alcohol Concentration (BAC) levels between 0.10 and 0.15 percent were only 40 to 50 percent. Sober failure rates (false-positives) ranged between four and eight percent (Sussman and Abernethy, 1973). TSC also conducted tests on an breath test interlock device (TSC Unpublished Report, 1973). The chemical sensor component of the device used a fuel cell as an oxidizer. These tests suggested to the researchers that the breath tester would not be practical as an in-vehicle device for two reasons. First, it was very susceptible to "cheating." The drunk driver could easily circumvent the device by either having a sober companion take the test for him or by using alcohol-free breath stored for this purpose in a balloon or air pump. Second, these tests showed that, due to naturally occurring chemical decomposition, the alcohol breath device would have to be inspected frequently, recalibrated, and/or replaced.

During this same period, Tennant and Thompson (1973), of General Motor's engineering staff, tested a Critical Tracking Task (CTT) for its discriminability in a series of laboratory tests. The CTT showed a great deal of promise, demonstrating failure rates of 50-75 percent at BAC levels of 0.10-0.14 percent and a very low sober failure (false-positive) rate of two to three percent. As the result of General Motor's work, and the development of other "second generation" behavioral testers, Oates (1973), under contract to NHTSA, conducted laboratory tests on four of the more promising of these devices. Included were the Critical Tracking Task, Complex Coordinator developed by the National Aeronautics and Space Administration (NASA) Langley Research Center, Quickey developed by Robert Smith, and Divided Attention Tester developed by TSC (based on earlier work by Moskowitz and Deprey, 1968).

Oates investigated the effects of different pass-fail criteria based on both individual and group performance, and different test scoring strategies, on the device's ability to discriminate between sober and alcohol impaired subjects. Further work (Oates; 1975a, b) suggested that two of the devices, Critical Tracking Task and Divided Attention Tester, were capable of high discriminability with BAC. Test failure rates of 60-90 percent were achieved at BAC levels of 0.15 percent. Different pass-fail criteria based on individual versus group performance scores did not result in significantly different pass-fail rates. The results of these tests were consistant with earlier tests by Tennant and Thompson (1973) and TSC.

Tests conducted by Oates suggested to NHTSA that the behavioral test concept was viable. The question of device circumvention remained unanswered, however. These tests indicated that circumvention of the devices required a proficient accomplice. Other questions remained unanswered also, e.g., what were some of the practical problems that would be encountered in implementing such a device. NHTSA decided that a controlled field test of the in-vehicle concept was necessary in order to answer these questions. The Critical Tracking Task (CTT) was selected by NHTSA for inclusion in a prototype in-vehicle system that would serve as the field test vehicle. That prototype in-vehicle system would be known as the Drunk Driving Warning System (DDWS). The CTT was selected for inclusion in the DDWS because its hardware development was more advanced than that of the Divided Attention Tester (DAT).

C. CRITICAL TRACKING TASK (CTT)

The Critical Tracking Task (CTT) was developed in the early 1960s to assess pilot and astronaut performance (Jex, et al., 1966). Since then it has proven to be an effective indicator of the effects of various environmental impairments in addition to alcohol impairment. The CTT requires the user to control an inherently unstable task not unlike that of balancing a broomstick on his/her fingertip. The user keeps the broomstick balanced by correcting for the instability. That is, as the stick starts to fall to the right the user corrects by moving his/her hand to the right; the stick then falls left, the user's hand moves left, etc. When used in a vehicle a needle display is mounted on the steering column. The testee uses the steering wheel to make left-right corrections. The CTT increases test difficulty by increasing the test instability (e.g., shortening the broomstick length) at a predetermined rate. As the test gets more unstable the user is forced to correct at a faster and faster rate (e.g., compare balancing a pencil as opposed to a broomstick). At some point the user's reaction time and eye-hand coordination cannot keep up with the broomstick and he/she loses control. This "critical instability" level or score is used to develop a pass-fail level for subsequent test trials on the device. The manner in which pass-fail levels are set and the number of test trials permitted prior to failing the CTT test are discussed in Section III. Pass-fail scores and testing strategies developed for use during field testing of the CTT when incorporated into the DDWS are discussed in Section IV.

D. DRUNK DRIVING WARNING SYSTEM (DDWS)

The Drunk Driving Warning System (DDWS), described in detail in Section VI, Appendix F, and by Peters, et al. (1975), is an in-vehicle system designed for use primarily with convicted drunk drivers. The DDWS requires that a short 10-30 second CTT steering competency test be successfully completed before the vehicle can be operated in a normal manner. If the CTT test is failed the driver must wait for a period of 10 minutes before attempting to repeat it. The DDWS permits the vehicle to be operated under conditions when the behavioral CTT test is not taken or is failed. However, when the vehicle is operated under such conditions, other drivers and the police are alerted by blinking emergency lights and by the horn honking intermittently at speeds of 10 mph and above.

The present DDWS consists of four basic components. The Critical Tracking Task (CTT) display located on the vehicle steering column is a meter with a centered indicator needle. The CTT test requires the driver to keep the needle in the center region of the meter by left-right movement of the steering wheel. The second component is an electronics module located in the trunk for activating the emergency blinkers and horn, and to store information on CTT performance (e.g., driver identification based on sitting weight, number of attempts made to pass the CTT, attempts to operate the vehicle when the CTT has not been attempted or has been failed). The third component is a data logger located in the electronics module. The data logger records all events associated with DDWS vehicle operations (e.g., ignition on-off, CTT performance, trips, sitting weight of driver). The fourth component is a seat-weight sensor installed in the front seat of the vehicle. The seat-weight sensor has two parts, serving both recording and anti-circumvention functions. The weight sensor measures the approximate sitting weight of the driver, which is transmitted to the electronics module in the trunk. The data logger will record, at the time of each trial, whether or not the weight falls within the predetermined range. The seat switch initiates recycling of the CTT if the driver moves off the seat and opens the vehicle door after having successfully passed the test. This prevents the CTT test from being taken by a sober accomplice.

SECTION III

CRITICAL TRACKING TASK (CTT) DATA ANALYSIS

A. OVERVIEW

One task of the current project involved a review and analysis of data collected on the Critical Tracking Task (CTT) during the mid 1960s and the 1970s. This was done to better understand the statistical properties of the CTT, thus providing a basis for developing strategies to optimize its sensitivity to alcohol impairment. Development of those strategies is discussed in Section IV. Two studies were reviewed from the mid 1960s which provided a comprehensive data base for determining the statistical properties of the CTT (Jex, et al., 1966; McDonnell and Jex, 1967). More recent studies on the CTT during the early 1970s provided data for optimizing CTT test strategies and pass-fail scores for use in this project (Allen and Jex, 1973; Peters, et al., 1975; Oates, 1973; Oates, et al., 1975a, 1975b).

Article B describes and reviews the various data bases mentioned above to reveal important sources of CTT performance variability. Article C reviews the effects of alcohol impairment on CTT performance. This latter analysis tells us something about the potential of the CTT for detecting alcohol impairment among convicted drunk drivers of the type to be tested later in the project with the DDWS. Article D summarizes conclusions drawn from our analysis of CTT performance data, providing a basis for developing the CTT test strategies and pass-fail criteria discussed in Section IV. Technical details of the data reanalysis task are given in Appendix A.

B. AVERAGE CTT PERFORMANCE

CTT scores were described in terms of deterministic and random effects. The deterministic effects involve what happens to average performance. Important deterministic effects considered were:

- Alcohol impairment, which acts to lower CTT scores.
- Trial order, or short-term CTT scoring trends due to warmup or fatigue effects that can lead to improvement or deterioration in scores from one test attempt to the next.
- Long-term CTT scoring trends due to learning with repeated task experience.

Alcohol impairment is of course the trend we are interested in detecting for any given subject. Given one or more trial scores we must decide whether or not a subject is drunk or otherwise impaired.

Considering data from four prior research studies over a number of subjects, we can see from Fig. 1 that a very consistent trend is obtained between BAC and percent decrease in CTT score. The knee of the curve occurs in the region of 0.04 percent BAC. By the time 0.10 percent BAC is reached the CTT score has degraded by approximately 9 percent. Beyond 0.10 percent BAC, CTT performance deteriorates rapidly.

The CTT is set up so that passing involves exceeding a preset criterion level. In order for the criterion to be valid for an individual, we would prefer that he/she exhibit a stable performance level, not influenced by short and/or long term trends. The Dunlap data (Oates, 1973; Oates, et al., 1975a, 1975b) were reanalyzed on this project and it was found that over six repeated trials minimal trial-to-trial effects were found. Thus, for a multiple-trial strategy where the subject is given several chances to pass the test, his chances of passing a fixed criterion score will be not much better or worse on the sixth try than on the first.

Both short and long term trends on CTT scores had been analyzed by Allen and Jex (1973) where four subjects were isolated in a space station simulator for three months. CTT scores were obtained two times a day for 12 weeks. Statistical analysis did not show any day-to-day trends throughout a week, but did show week-to-week improvement over the course of the 12 week isolation period. This trend indicates that over a long period of time pass criterion levels would have to be slowly elevated in order to minimize the chance of the drunk driver passing the

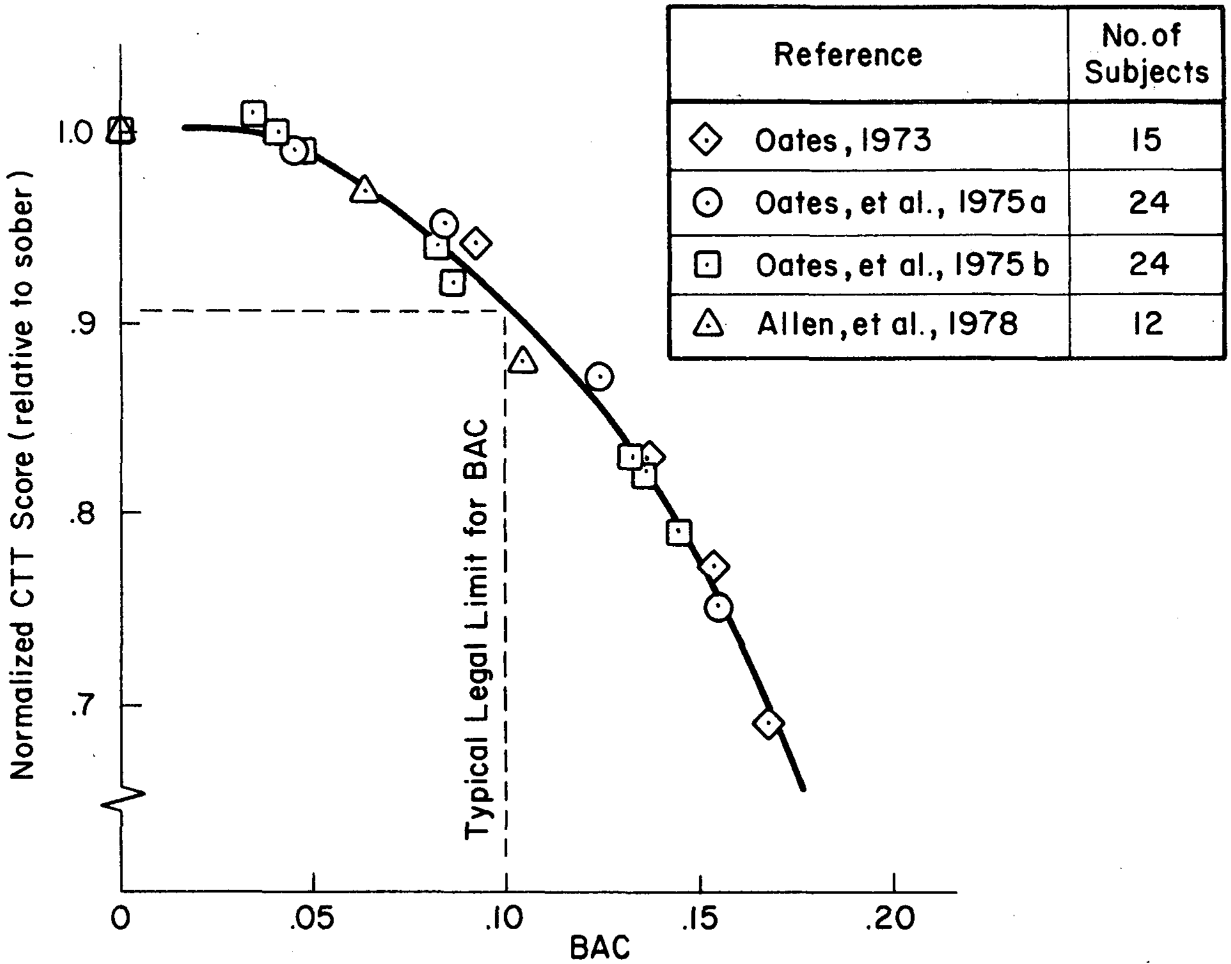


Figure 1. Effect of BAC on Normalized CTT Score

test. A feature could be programmed that would automatically elevate pass levels to keep up with long-term improvement trends, but this feature was not included in the DDWS equipment tested in this project.

C. VARIABILITY IN CTT PERFORMANCE

Random effects on CTT performance must also be accounted for in order to properly discriminate between sober and drunk drivers. Random effects lead to variations in performance without any apparent causal factors. The important sources are:

- <u>Run-to-run</u> -- random score variations from one run to the next for a given subject.
- <u>Subject-to-subject</u> -- random average performance differences between randomly selected subjects.
- <u>Subject response-to-alcohol</u> -- random differences in sensitivity to alcohol between randomly selected subjects.

The effects of run-to-run variability can be minimized by the use of multi-trial strategies as discussed in the next section, such as taking the average of n trials or allowing a maximum of n trials in which to achieve a passing score. Subject-to-subject variability is accounted for by setting individual pass criteria for each subject. Accounting for individual subject sensitivity to alcohol requires establishing each subject's tolerance, which is not practical. However, if we accept the hypothesis that critical task score is a measure of impairment rather than BAC, and that impairment is the important effect, then the issue of alcohol sensitivity is by definition not critical to test effectiveness.

Now let us consider the relative importance of the above three sources of variability. In Table 1 we have summarized the results of analyses conducted on CTT score data bases, including four alcohol studies, and three experiments involving other types of behavioral impairment (e.g., isolation, auditory noise, ship motion). Here we see that run-to-run and subject-to-subject variability have equal standard deviations of 8.6% of the sober mean CTT score when averaged across all the Table 1 experiments. These results are not much different from the

TABLE 1. COMPONENT VARIABILITY SUMMARY FOR CTT DATA. STANDARD DEVIATION EXPRESSED AS A PERCENTAGE OF SOBER MEAN SCORE. DATA OBTAINED THROUGH ANALYSIS OF VARIANCE PROCEDURES

REFERENCE	EXPERIMENT	NO. OF SUBJECTS	MEAN CTT SCORE	CTT VARIABILITY COMPONENTS (Percent)		
				RUN-TO-RUN	SUBJECT-TO-SUBJECT	SUBJECT × STRESS INTERACTION
ALCOHOL STRESS						
Oates, 1973	Dunlap: Second Generation ASIS	15	4.75	8.9	9.2	6.2
Oates, et al., 1975a	Dunlap: ASIS Lab Test Phase I	24	4.64	9.4	10.3	5.7
Oates, et al., 1975b	Dunlap: ASIS Lab Test Phase II	24	4.62	9.6	9.3	4.3
Allen, et al., 1978	STI: Alcohol and Marihuana	12	4.70	6.8	7.0	6.0
OTHER STRESSES						
Allen and Jex, 1973	90 Day Confinement	9	6.5	6.2	4.9	5.1
Allen, et al., 1975	Broadband Noise	4	3.1	9.1	5.8	3.9
Jex, et al., 1977	Ship Motion	8	5.14	10.0	14.0	2.7
OVERALL AVERAGE VARIABILITY				8.6	8.6	4.8

large body of behavioral task literature where between-subject variability is always significant and many times is the largest source of variability.

D. SUMMARY AND CONCLUSIONS

Referring back to Fig. 1, recall that 0.10 BAC gives approximately a 9 percent decrease in average CTT score. Here we see that run-to-run and subject-to-subject variabilities give about the same random effect in score levels. The conclusion to be drawn from these results is that both sources of variability must be treated in order to maximize CTT discriminability between sober and drunk drivers. As mentioned previously, the solution to variability between subjects is to use individualized pass criteria. Otherwise, given the relative magnitude of the variability components analyzed here, a universal pass criterion would lead to great disparity between subjects in their ability to pass the test when sober or drunk. The solution to minimizing run-to-run variability is to employ multiple trial strategies, as discussed in the next section.

SECTION IV

CTT TEST OPTIMIZATION

A. OVERVIEW

A task on this project was devoted to the development of optimum test strategies and pass-fail criteria for the CTT in order to maximize the discrimination between sober and alcohol impaired subjects typical of the type to be assigned the Drunk Driving Warning System (DDWS) later in the project. This task also involved a laboratory validation experiment carried out to verify the adequacy of CTT test strategies and pass-fail criteria established earlier in the task. This section of the report describes test strategies and pass-fail criteria development. Section V describes the laboratory validation experiment.

The basic issue relative to CTT discriminability is how to reach reliable decisions concerning a driver's alcohol impairment level within a minimum number of test trials. From a statistical decision theory perspective we are confronted by two types of error. The first type of error occurs when a <u>sober driver fails</u> to pass the test. The second type of error occurs when a <u>drunk driver passes</u> the test, thus creating a safety hazard.

The technical details for optimizing CTT test discriminability are given in Appendix B and are summarized below in Article B. Considerations for setting CTT pass-fail scores are discussed in Article C. Article D discusses procedures for achieving stable CTT learning patterns and for estimating individual pass-fail criteria. Article E summarizes and draws conclusions about optimum CTT test strategies, pass-fail criteria and learning protocols.

B. TEST STRATEGY

Minimizing the effect of within-subject or run-to-run variability in CTT test scores requires the use of multiple-trial test strategies. For example, the average score of two trials will be a more reliable estimate of the subject's performance capability than either one of the trial scores alone. Three strategies of practical interest were analyzed in this study.

- Average of n trials -- the classical means for obtaining reliable estimates of random variables. The variability of the estimate increases with the number of samples. For any sizable sample this strategy is inconvenient, however, because all n trials must be taken.

- One out of n trials -- the subject is allowed n attempts to reach the pass level. Since the subject can pass on any given trial, this strategy may lead to a low number of trials for the sober driver, which is a matter of convenience.

- Sequential selection -- the subject is given a pass level and a lower fail level. On any given trial the subject may pass or fail, or be allowed another trial if his/her score is in between the pass and fail levels. This strategy can also lead to a quick decision (i.e., few number of trials).

The above strategies were applied to the Dunlap data bases Oates, 1973; Oates, et al., 1975a, 1975b) discussed previously. As discussed in more detail in Appendix B, the one-of-n strategy was determined to be the best practical strategy for the following reasons:

- The sober driver can pass the test with a low average number of trials (i.e., 1.7 trials).

- Individual pass levels are near the subjects' average performance levels, which leads to fairly stable decisions (i.e., not dependent on low probability good or bad performance).

- Subjects' pass scores can be reliably estimated because they are near average performance levels.

Results for three values of n are illustrated in Fig. 2. Here individual pass levels were selected for each subject to give a relatively low chance of sober failure, i.e., 2.5 percent. As BAC is increased the failure rate gradually increases up to the region of 0.07 percent or so, where the failure rate climbs steeply. For four attempts, the failure rate reaches 50 percent in the region of 0.11 percent BAC. At 0.15 percent BAC the test failure rate is over 80 percent.

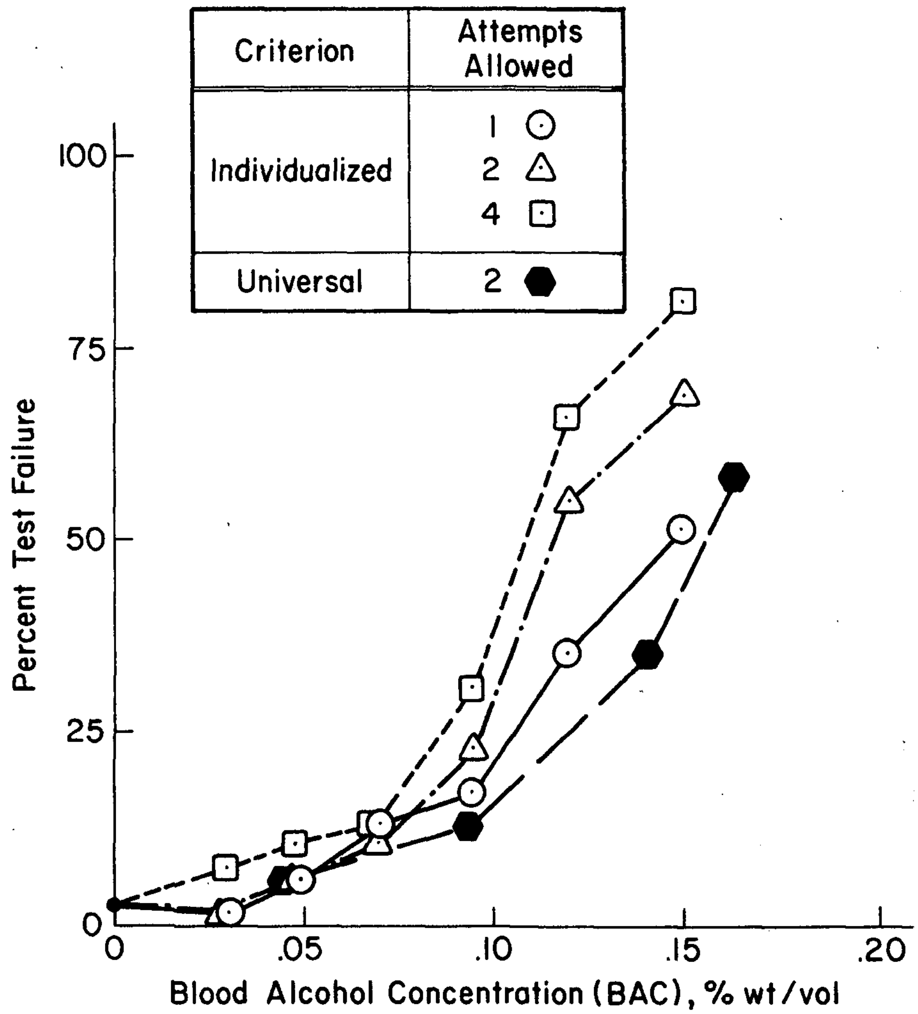


Figure 2. Effects of Alcohol Intoxication on DDWS Test Failure for one out of n Test Strategies

Note also in Fig. 2 a plot for data analyzed with a single universal criterion for all subjects. The universal criterion discriminability curve starts at the same sober failure rate as the other curves, but is much less sensitive to BAC (e.g., in the region of 0.11 percent BAC the failure rate with individual criteria is double the universal criterion data). Because of performance differences between subjects with a universal pass criterion, some undoubtedly fail almost every trial, while others hardly fail any.

Because between-subject variability is quite high (as is true of most behavioral tasks), the fallacy of a universal criterion lies in its effect on subjects who are not near the average (i.e., the good performers and poor performers). The key to improved discriminability with individual pass criteria is to get a reliable estimate of each subject's well-trained performance capability, then set the pass level according to statistical decision theory principles. The means for accomplishing this is discussed below.

C. PASS CRITERIA

When used operationally, the DDWS requires that the driver pass the CTT performance test. In setting pass criteria we must take into account both statistical and behavioral considerations. Each subject's pass criterion must be determined by procedures applied to performance data that accurately reflect his/her performance ability. From a statistical point of view the specific procedures depend on the details of the test strategy and are determined by statistical decision theory as discussed in Appendix B. The behavioral problem concerns obtaining a set of performance data that is representative of the subject's well-trained behavior. The training problem is summarized in Article D.

A subject's chance of passing any given trial depends on the pass criterion level and the variability inherent in the CTT score from one trial to the next (i.e., the between-trial variability). Assume a subject can exceed a given score 50 percent of the time (i.e., his average or median score). If we set this score as a pass level, then his chances of passing any trial are 50 percent. Now allow the subject more than one chance to pass the test. From statistical theory, the chance of failing to reach a 50 percent pass level on both trials is 25 percent (i.e., for independent trials, the probability of failing both is $0.5 \times 0.5 = 0.25$). If three trials are allowed, then the chance of failing all three trials is 12.5 percent; for four trials it is 6.25 percent, etc.

Now let us turn the problem around and state that the subject will be allowed four trials in which to exceed the pass level once, and in addition we want to set the probability of his failing four trials in a row (i.e., failing the test and incurring a 10 minute wait) to be relatively low. Of course, the issue of exactly what level to set the sober failure rate at is a matter of judgement. For the purposes of this project, it was the consensus of the investigators and NHTSA personnel that a 2.5 percent failure rate would be acceptable to drivers, and still give adequate BAC discriminability. Note, however, in the Appendix C analyses that higher BAC discriminability can be achieved by increasing the sober failure rate, and future DDWS applications may wish to take advantage of this option.

When the desired sober failure rate is 2.5 percent, then the chance of the subject failing any one trial when sober should be 40 percent (i.e., the probability of test failure given 4 trials is $0.4 \times 0.4 \times 0.4 \times 0.4 \cong .025$). This result says that the subject's pass criterion level should be set at slightly below his/her average or median performance capability. Typical average CTT scores are in the range of 4.5 to 5.0 units, and analysis of the Dunlap CTT data (Oates, 1973; Oates, et al., 1975a, 1975b) shows that the 40 percent failure level is given approximately by taking the average score and subtracting one-tenth (0.10) of a unit (Appendix B).

The above results have several desirable attributes. First of all, since the desired pass level for a given subject is near his/her average performance level, most trial attempts end up with scores in this region. Thus the pass level for a given subject can be determined relatively reliably given a set of performance scores. Secondly, from the subject's point of view the test can be passed by reaching a near-average performance level so that routine sober operation does not represent excessive stress. Finally, statistical theory and analysis of the Dunlap data shows that the sober subject can pass the test on the first trial about 60 percent of the time, and within two trials about 85 percent of the time. These last two attributes might be interpreted from the point of view of subject convenience, but they should also be considered in terms of traffic safety. Subject experience in taking the test should not induce excessive stress or anxiety that might impair subsequent driving behavior.

Now, given a set of CTT performance data for a given subject, consider how the pass criterion level might be selected. Since the level is near his/her average level, an individual's mean performance could be computed by standard procedures. The pass level could then be obtained by subtracting a tenth of a unit (0.10) as discussed above. The problem with this is that a subject's first pass level will be determined from training data which might include low scores that are influenced by incomplete learning behavior. Subsequent to training low scores can also arise due to a myriad of events, including occasional impaired behavior in the unsupervised field setting.

In order to minimize low score artifacts, the CTT performance scores can be organized in ascending order so that the low scores can easily be identified and ignored. Then the middle and top scores can be used to set the pass level. As described in Appendix E, several methods were analyzed for obtaining accurate estimates of an individual's pass level. The most consistent method is to determine the 40th percentile score (i.e., 40 percent of CTT scores below this value) from a cumulative percentage probability plot of the data. As discussed in Appendices C and E, this can be accomplished by determining the median performance level, and setting the pass level at 0.1 unit lower. Since this technique depends on determining a subjects average performance level, it can give reliable pass level estimates based on relatively few trials (say 30 or more). This procedure is consistent with obtaining the first pass level for an individual at the conclusion of training, and also provides a simple means for subsequently upgrading the pass level based on unsupervised field test data in order to minimize long-term learning.

D. TRAINING

Two objectives are satisfied with training. One is to obtain a set of CTT performance scores that represent a subject's well-learned performance ability so that his/her pass level can be reliably estimated. The second objective is to give the subject adequate experience with the CTT so that he/she can routinely pass the test when sober with criterion levels that are still stringent enough to deter drunk driving. From a practical point of view, we also want to identify procedures that will minimize training logistics. Details on CTT training experience and procedures are given in Appendix E; a summary follows.

In the training literature (Hovland, 1951; Woodworth and Schlosberg, 1964) the issues of concentrated vs. distributed practice are identified as influencing learning progress. In essence, concentrated practice occurs over a relatively short period of time, while distributed practice occurs over several encounters. In general, learning over any one concentrated encounter reaches some limit, and performance may even deteriorate past some point in a lengthy exposure (e.g., due to fatigue, saturation, etc.). During a subsequent exposure subjects may start off at a level greater than the best previous performance, and continue to learn up to some new saturation level.

Past CTT research has shown that there are short- and long-term learning trends (Allen and Jex, 1973). The objective of CTT training is to get past the short-term phase of learning. Analysis of the Dunlap data shows this can be accomplished within three separate training sessions, incorporating 100 trials per session (Appendix E). With this amount of practice, the better scores in the last session appear to allow a relatively reliable estimate of a given individual's CTT pass level.

Another important training issue concerns the motivation or incentives provided for inducing rapid learning. Past experience with the

CTT and other psychomotor tasks has shown that subjects who come to an experiment with positive attitudes (e.g., volunteers, professionals such as pilots) can be motivated with positive incentives such as monetary rewards and competition for high scores. Subsequent experience on this project with court-assigned subjects (i.e., DWIs) showed that positive incentives did not reliably motivate all subjects. Because of this experience, a negative incentive consisting of a forced wait time for test failure was developed as discussed in Appendix E. This procedure appears to consistently motivate all subjects to pass the test as often as possible.

E. SUMMARY AND CONCLUSIONS

Reanalysis of previous CTT data bases has shown that to minimize the two types of decision errors the best procedure is to use:

- One pass out of 4 attempts
- Individualized pass level criterion
- Pass criterion set at the subject's 40 percent single-trial failure level
- Three separate training sessions of 100 trials each

Additionally, a forced wait time for trial failure motivates the subject such that complete learning is achieved and his/her pass level can be reliably estimated.

SECTION V

CTT LABORATORY VALIDATION EXPERIMENT

A. EXPERIMENTAL DESIGN

A CTT laboratory validation experiment was conducted to verify the adequacy of CTT test strategy and pass-fail criteria setting procedures developed from analyses reported in Sections III and IV of this report. In addition, the experiment, discussed below, was designed to investigate three related questions:

> Long Term CTT Learning -- can learning trends be inhibited by stopping the test immediately upon achievement of the pass-fail criteria?
>
> Learning While Alcohol Impaired -- does experience with the CTT under conditions of alcohol impairment improve subsequent performance under alcohol?
>
> CTT Performance Impairment -- is CTT pass-fail performance under sober and alcohol impaired conditions consistent with performance under similar sober/alcohol conditions on a driving simulator task?

The validation experiment design, in Table 2, satisfied these objectives. Each subject (convicted DWIs) participated in three training sessions and four experimental sessions in order to allow consideration of short- and long-term learning trends. The first three experimental sessions included a placebo and two alcohol sessions in order to permit detection of learning under alcohol. The 24 subjects were subdivided into two main groups to test whether stopping the CTT at the pass level would inhibit long-term learning. The two groups were Test-to-Limit (TTL) and Test-to-Pass (TTP). In the TTL condition each trial is terminated only when the subject loses control of the CTT. In the TTP condition the trial is ended when the subject loses control or reaches his/her predetermined pass level -- whichever happens first. Each main group was further subdivided into three subgroups so that the placebo experimental session was encountered first, second, or last by equal numbers of subjects in order to avoid biasing any of the results due to learning trends. Each subject subgroup was balanced for age, with one

TABLE 2. EXPERIMENTAL DESIGN

Training Design For All Subjects

1st Session	25 Blocks 4 Trials/Block
2nd Session	25 Blocks 4 Trials/Block
3rd Session	25 Blocks 4 Trials/Block
4th Session For Subjects With Motivation Problems	25 Blocks 4 Trials/Block

ALL TRAINING TRIALS WERE TEST TO LIMIT

Experimental Design for Formal Treatment Conditions

Session	SUBJECT GROUP					
	Test-to-Limit			Test-to-Pass		
	1	2	3	1	2	3
1	P	A	A	P	A	A
2	A	P	A	A	P	A
3	A	A	P	A	A	P
4	Final short session-sober baseline "test-to-limit" block					

P = Placebo For Each Group N = 4

A = Alcohol

subject from each of four age categories. These categories were 21-25, 26-30, 31-40, and over 40.

A short sober baseline session was given to all subjects several days after their last experimental session. They were given 12 trials on the CTT under test-to-limit conditions. This took about 1/2 hour and was required in order to evaluate learning differences between the TTL and TTP groups.

B. PROCEDURES

1. Subjects

Subjects were convicted drunk drivers obtained through the cooperation of the Los Angeles Municipal Courts. Twenty-four so-called volunteers were permitted to participate in the experiment as a condition of probation, and, in exchange, received a reduction on their fine. The Minnesota Multiphasic Personality Inventory (MMPI) was administered as a screening procedure. If the three validity scales showed the profile to be unreliable or if the individual scored over the 70th percentile on any of the eight clinical scales, he/she was eliminated. Persons with felony criminal records were also eliminated.

2. Training

During each of the three training sessions each subject was required to complete 100 trials on the CTT. The pass criterion at the outset was 2.9 and all trials were conducted in the test to limit mode. The 100 trials were divided into 25 blocks of 4 trials each. If 4 out of 4 trials in a block were passed, the criterion was raised 0.2. If 3 out of 4 trials were passed the criterion was raised 0.1. The subject's feedback consisted of either a green "pass" light or a red "fail" depending on whether or not the pass criterion was exceeded. The block means of the third session were computed. The second highest block mean, reduced by 0.3, provided the pass criterion for that subject's experimental sessions. Previous analysis had shown this procedure to give appropriate pass levels in order to achieve a 2.5 percent sober failure rate (see Appendix E).

At the first training session the subject was also given a brief orientation to the driver simulator and the concept of a driving scenario. Following this, he "drove" a practice run. During this run, the experimenter verbally explained each maneuver, and the maneuver was practiced until it could be performed fairly consistently.

Following this initial orientation the subject drove two computer-controlled driving scenarios. These scenarios were similar in both length and complexity to those used in the actual experiments. At the second and third training sessions, the subjects drove two scenario runs. They were allowed approximately 5 minutes of "warm-up" prior to driving the first run. During all three sessions the first drive was considered practice and the second was driven using the reward-penalty structure. This served to motivate the subject to improve his performance and strive for optimum performance. For a detailed description of the driving scenarios and the reward-penalty structure see Appendix D.

3. Experimental Sessions

The formal experimental test sessions began with a breath test using a gas chromatograph intoximeter. If the subjects tested sober they were given a baseline CTT test block and a driving simulator run. The subjects then consumed three drinks calculated to bring them up to a BAC of 0.15 percent. The drinks consisted of a measured amount of hard liquor (e.g., vodka, bourbon, etc.), based on body weight, and mixer (e.g., orange juice). Mixer was required in order to prepare credible placebos. Placebos were prepared with a small amount of liquor floated on top of the subject's preferred mixer.

Next, the subjects were administered a CTT performance test block at 0.15 BAC and given a meal, followed by CTT performance test blocks at 0.10 and 0.075 BAC during the sobering up phase. The CTT test blocks involved three groups of four trials each separated by ten minutes each to simulate the DDWS function which allows a test retake 10 minutes following a prior test failure. Driver simulator runs were conducted following the CTT test blocks. During one alcohol session the simulator runs were administered at the 0.15 and 0.10 BAC levels. During the other alcohol session, the simulator runs were administered at the 0.10 and 0.075 BAC levels.

C. RESULTS

CTT learning proceeded at a consistent rate for most subjects as evidenced by Fig. 3. As stated previously, several subjects were unmotivated and performed poorly. We resolved this problem by requiring a fourth training session of those individuals. Since then, we have employed a time penalty technique that motivates performance such that complete training is accomplished in three sessions. This procedure is explained in Appendix E.

Analysis of between groups effects showed that the Test-to-Limit (TTL) group did not score much better than the Test-to-Pass (TTP) group. Thus the Test-to-Pass configuration does not appear to inhibit learning over the time span considered in this laboratory experiment. Also, no statistically consistent performance trends over sessions were noted between the three session order subject groups indicated in Table 2. Thus, CTT experience under alcohol impairment does not appear to improve subsequent alcohol impaired performance.

The averaged data were compared with data from prior experiments (Tennant and Thompson, 1973; Oates, 1975a, 1975b), and were found to be consistent in terms of both average performance and variability. The relatively low variability (about 9 percent) is consistent with all the other CTT experiments, as discussed previously. The discriminability results agreed with statistical model analysis discussed in Appendix C, as Fig. 4 shows. At the lower BAC levels the actual tests were slightly more sensitive than the model predictions.

While these results prove that the one-of-n strategy CTT Failure Rate is a reliable correlate of BAC, the more important result is the correlation between driving simulator performance, and both CTT failure rate and BAC. Figure 5 makes this comparison for both accidents and speed exceedances. In both figures, CTT Failure Rate (one-of-four pass strategy) is shown by the solid hexagonal symbol, and simulator driving

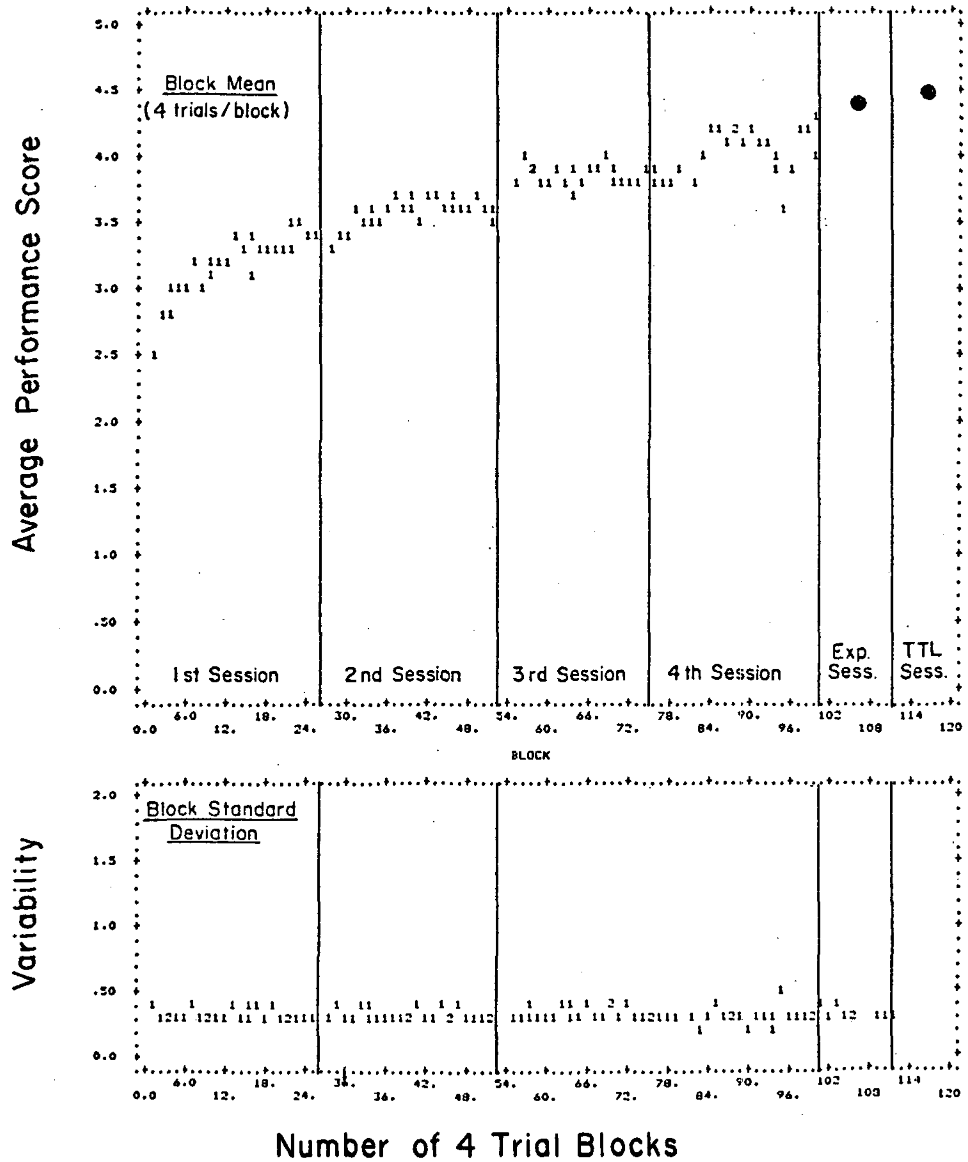


Figure 3. CTT Learning Curve Averaged Across All Subjects

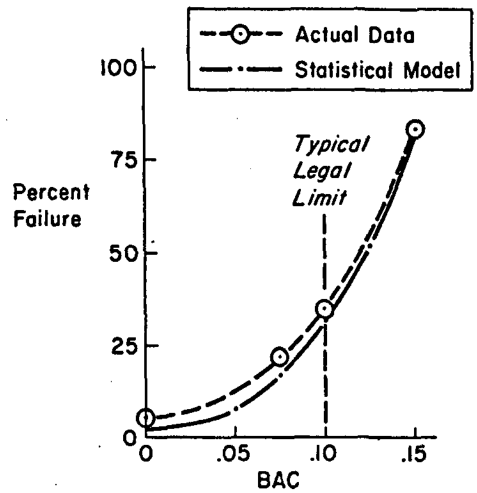


Figure 4. CTT Alcohol Impairment Discriminability Validation Data Compared with Statistical Model Prediction (One Pass Out of Four Tries Strategy)

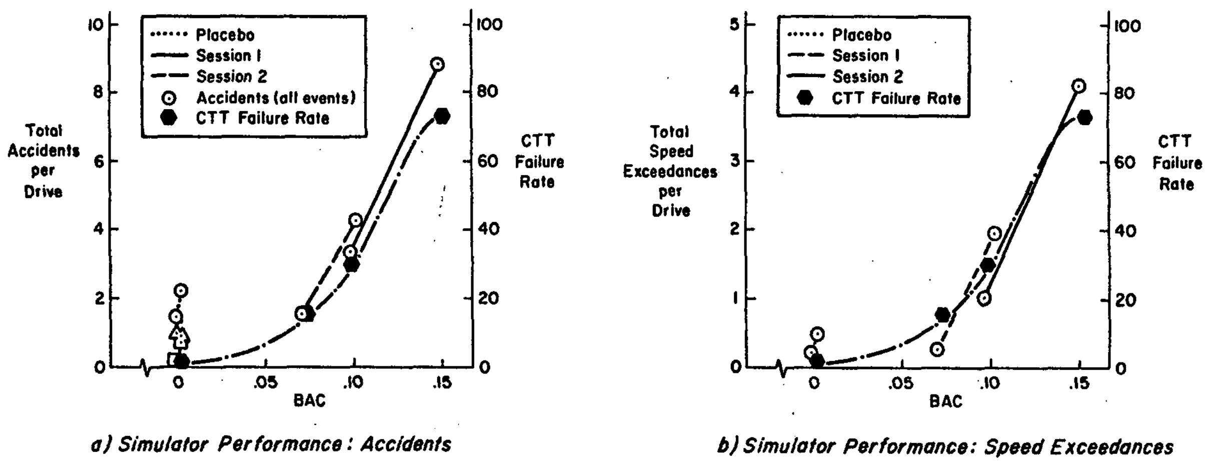


Figure 5. Comparison of Driving Simulator Performance CTT Failure Rate

performance by the open symbols. As CTT Failure rate increases, the actual driving performance suffers at almost the same rate implying an excellent correlation between CTT test failure rate and simulator driving impairment.

The above results show comparable trends between CTT and simulator performance. In Fig. 6 the BAC sensitivity of CTT and simulator performance are also noted to be consistent with real world accident rates. In Fig. 6, the relative probability of accidents as a function of BAC is plotted for several real world accident studies. Note that similar trends occur for CTT discriminability and simulator and real world accident rates in that performance begins to deteriorate rapidly in the region of 0.08 BAC.

In order to determine the direct correlation between CTT and simulator performance, accidents and speed limit exceedances were counted according to whether they were associated with a CTT pass or failure. The results are given in Fig. 7. Here we see that the accidents associated with passing the CTT stay relatively constant with BAC, while accidents associated with CTT failures increase dramatically with increasing BAC. A similar, although less consistent, trend is noted for speed limit exceedances.

These correlations between predicted and actual test performance show that it is now possible to both predict and verify vehicle operator impairment using a psychomotor test such as the Critical Tracking Task.

D. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Analysis of past CTT data bases shows that score variability is almost equipartitioned between run-to-run and subject-to-subject components. Furthermore these variability components (i.e., standard deviations) are approximately the same magnitude as the score decrement due to an alcohol impairment of 0.10 BAC. Thus, in order to achieve any degree of acceptable discriminability, the consequences of the variability sources must be dealt with as follows:

- Run-to-run variability -- use a multiple trial test strategy

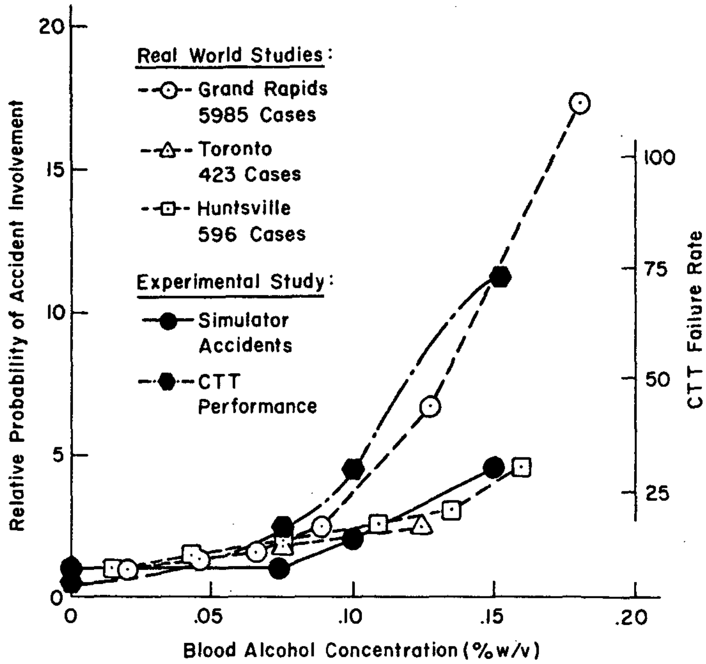


Figure 6. Comparison of CTT and Simulator Performance with Real World Accident Rates (Real World Data Adapted From Hurst, 1973 and Voas, 1974)

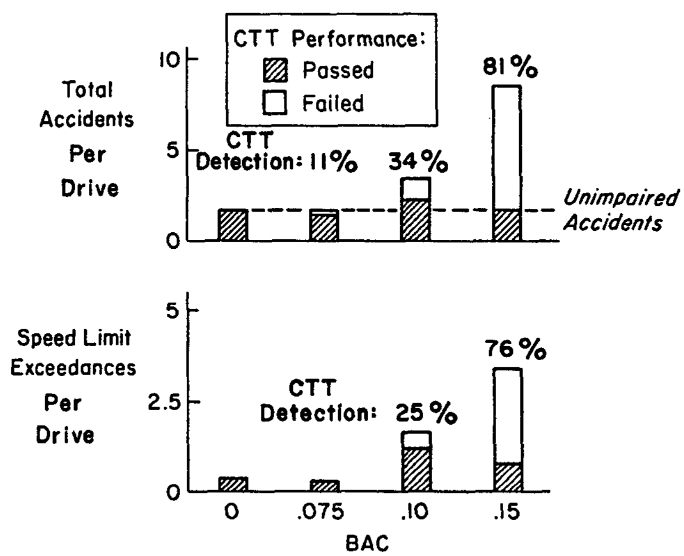


Figure 7. CTT Discriminability of Impaired Driving Simulator Performance

* Subject-to-subject variability -- set individual pass levels for each subject

Test strategy should be analyzed in terms of minimizing sober failures and maximizing drunk failures. In addition, convenience to the sober driver should be considered. Procedures for setting individualized pass levels depend on a given strategy and the tolerable sober failure rate. Statistical decision theory is used to determine the exact relationships between sober failure rate and pass level for a given strategy. Using these procedures, much improved test discriminability is shown over past analysis efforts (Oates, et al., 1975b).

Estimating individual pass levels requires analyzing the cumulative percentage distribution of a given subject's CTT scores. This is initially done for data from the final training session, and pass levels can be updated subsequently in the same manner. Multiple training sessions must be employed, and adequate trials given within each session to achieve maximum within-session learning.

CTT pass-fail performance under sober and alcohol-impaired conditions is consistent with performance on a driver simulator task under the same sober and alcohol-impaired conditions. These tests were run with DDWS procedures and parameters set at optimum values, and the results strongly support and validate the recommended field test conditions summarized below.

E. FIELD TEST RECOMMENDATIONS

Based on analysis and results of the validation experiment, the following recommendations are made for conducting the DDWS field trials:

* <u>Test strategy</u> -- four attempts to exceed the pass level.
* <u>Pass level</u> -- set individually for each subject, based on the single-trial 40th percentile score level. Initial level set from last training session, subsequent updates based on check-in data.
* <u>Training</u> -- three sessions of 100 trials each, use wait-time penalties for test failure to motivate good performance.

SECTION VI

DRUNK DRIVING WARNING SYSTEM (DDWS) SYSTEM CONFIGURATION

A. OVERALL INSTALLATION

The DDWS in its current configuration is installed in eleven 1978 Chevrolet Novas. There are four basic components to the system: 1) the Critical Tracking Task (CTT) display unit and steering sensor, 2) an electronics module located in the trunk which provides all computations and activates the hazard lights and horn, 3) a seat-weight sensor, and 4) a data logger that records test trials and driving events. A technical system description is given in Appendix F. A summary description is as follows.

The CTT display unit and steering sensor are located adjacent to the vehicle steering wheel, as shown in Fig. 7. The display consists of a meter with a centered indicator needle; the steering sensor is comprised of a potentiometer driven by a gear mounted on the steering wheel. The test requires the driver to keep the needle in the center region of the meter by appropriate movement of the steering wheel, while the trunk-mounted CTT computer causes it to fall to one side or the other. The rate of needle divergence is increased automatically by the CTT until the driver loses control of the task. The rate of needle divergence commanded by the CTT is the measure of task difficulty, and the difficulty at the point of control loss is taken as the performance score. Higher scores indicate better performance.

The electronics module located in the trunk (Fig. 8) contains the CTT computer and decision logic to activate the various options of the system. There are 7 different strategy choices. This allows the experimenter to choose how many passes (1 or 2) are required out of how many attempts (1-4). The experimenter also sets the appropriate pass level and test mode, which controls the test termination criteria. In the test-to-limit mode the test never is terminated until the subject

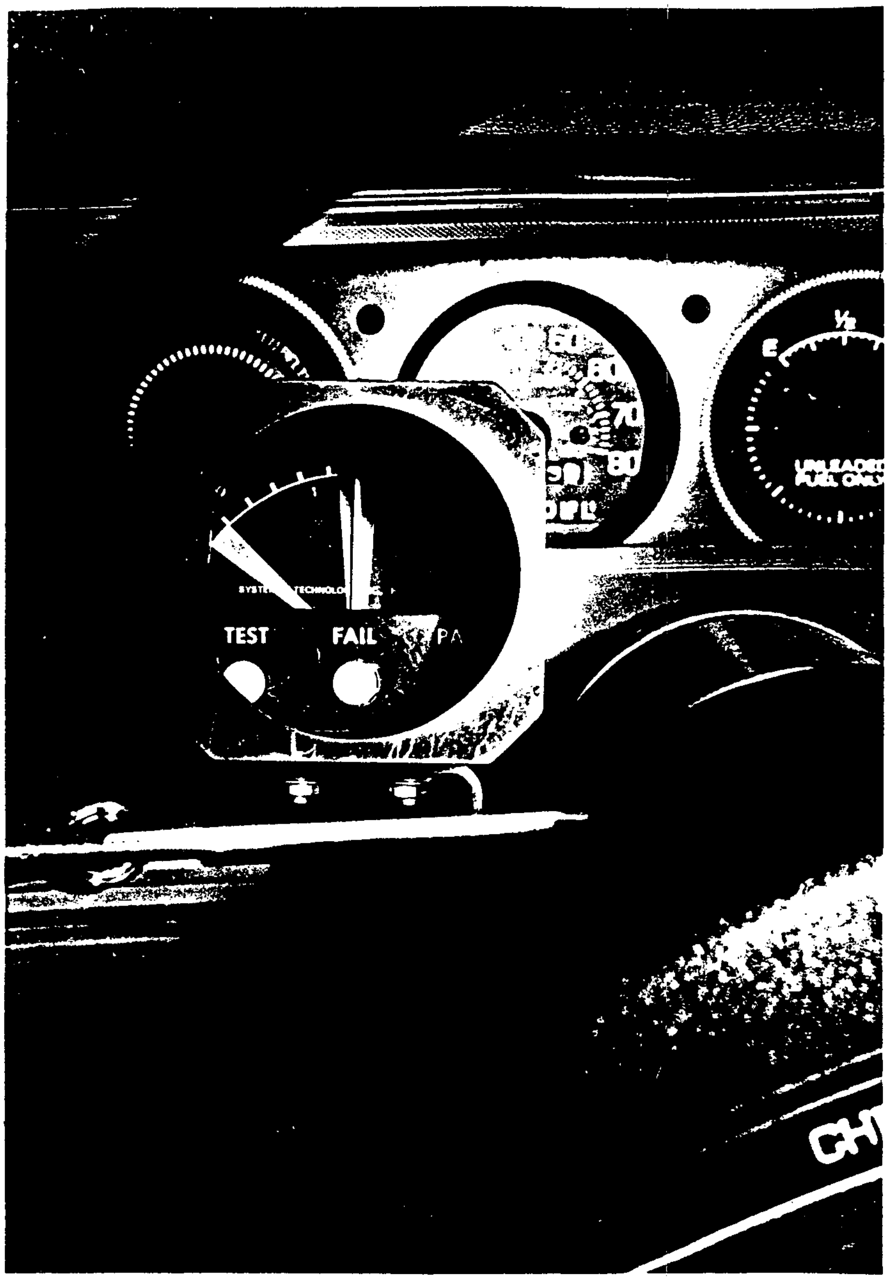


Figure 8. Dash Mounted CTT Meter

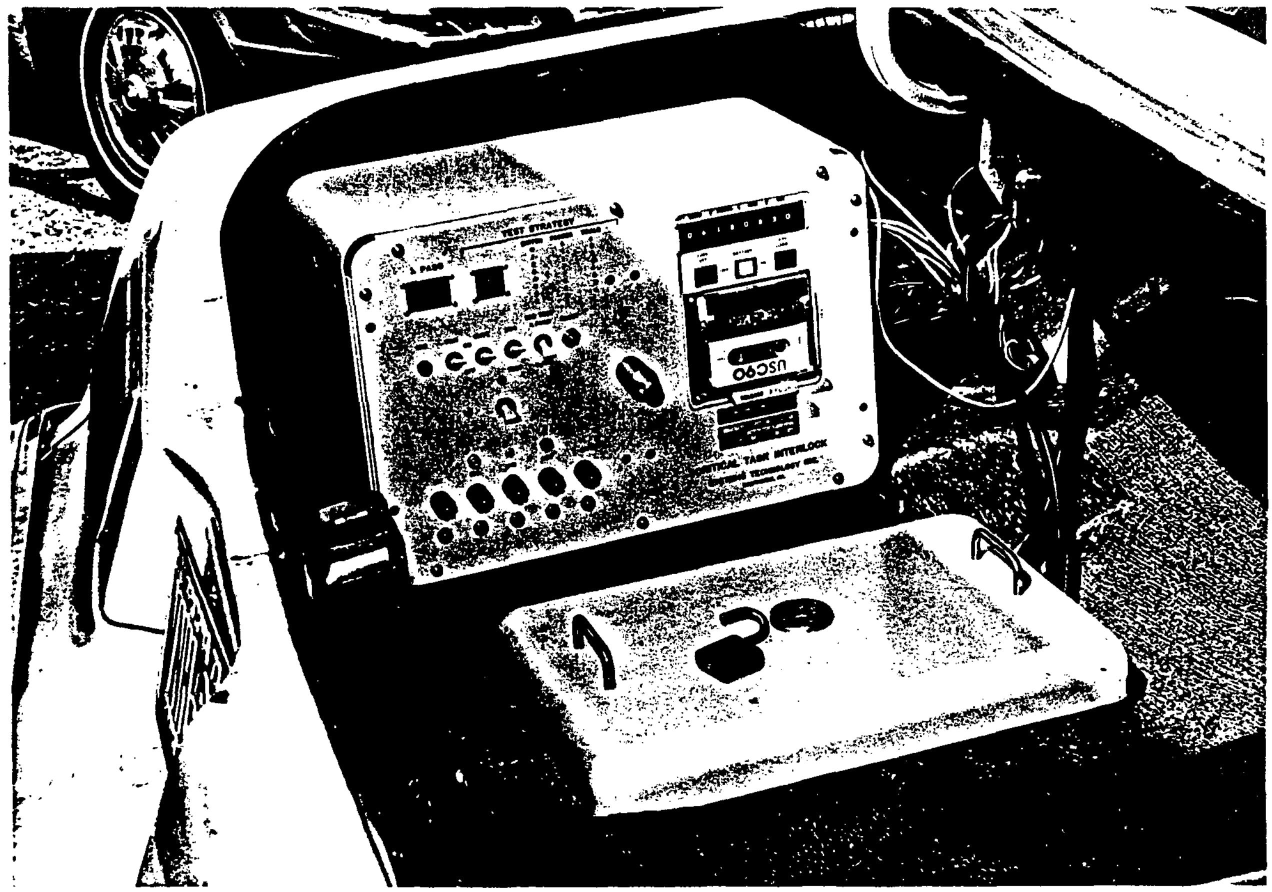


Figure 9. Trunk Mounted Electronics Module and Data Logger

loses control, even though the pass level has been exceeded. In the test-to-pass mode the test is always terminated when the pass level is reached, in an attempt to minimize long-term learning. The trunk module also contains the mechanism by which the alarms are deactivated upon successful completion of the test.

The third component of the DDWS system is the seat weight sensor. The seat weight sensor has two parts — a seat switch and a weight sensor. This system serves both recording and DDWS anti-circumvention functions. The weight sensor is a strain gauge installed under the driver's seat. It measures the driver's weight and weight distribution. At the time of each trial the data logger records whether or not the driver's weight distribution falls within a predetermined range. The seat switch signals to the electronics module when the driver leaves and enters the car. This activates the system to require a new test each time the driver enters the car and prevents the driver from switching with someone else after the test is passed. The components of the seat weight sensor system are shown in Fig. 9.

The fourth component of the DDWS system is the data logger. There are four different events that are recorded on the magnetic recording tape. These events are: 1) ignition turned on, 2) ignition turned off, 3) test trial score, and 4) driving over 10 mph without passing the test. Each event is recorded with the date and time of the event. Additionally, when a test trial is recorded, it is also noted on the tape whether the driver's weight and body configuration matches the one set up in the computer.

B. BASIC OPERATION

The DDWS requires the driver to pass a brief (15-40 seconds) CTT steering competency test before driving the car in a normal manner. The test must be passed in order to deactivate alarms consisting of the emergency flasher system and the horn. Because DDWS is a warning system, and does not prevent the vehicle from running, the car can be driven without passing the test. However, if the test is not passed the emergency flashers operate, and if the car is driven above 10 mph the

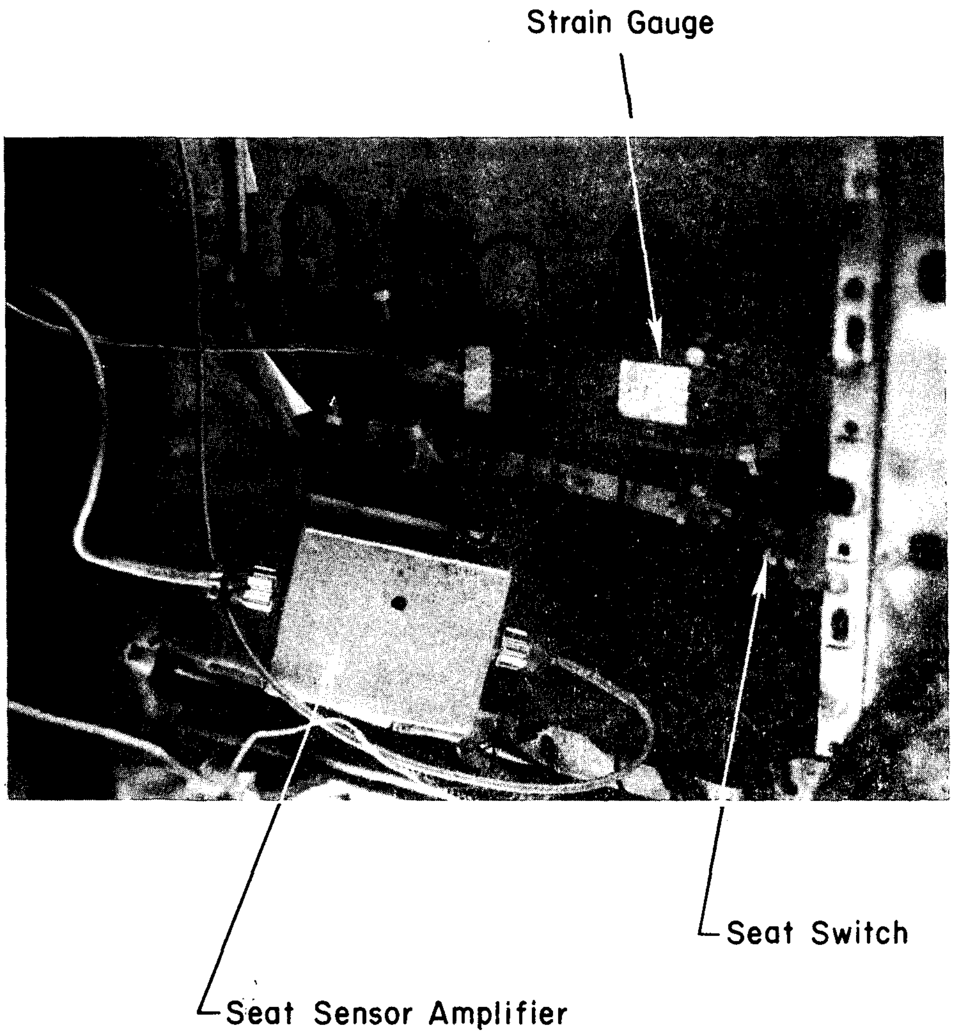


Figure 10. The Seat Weight Sensor System
(Underside view of Driver Seat)

horn honks at one second intervals. If the test is failed the driver must wait 10 minutes before retesting is permitted.

When the driver is seated in the driver seat of the DDWS vehicle and the ignition is turned on, a YELLOW test light located in the CTT steering column display is activated. The driver then presses a test button located on the left-hand side of the steering column to initiate the test trial. He/she then tries to keep the display needle centered using the steering wheel (see Fig. 7), while the CTT causes the needle to fall from side to side. The rate of needle divergence is increased automatically by the CTT until the driver loses control of the task, or achieves a pre-determined pass level. Each driver is given an individualized pass level based on his/her sober performance capability as described in Section IV. If the driver achieves his/her pass level, a GREEN light located in the CTT steering column display is activated, and the vehicle can be operated in a normal manner. If the driver loses control of the task before his/her pass level is achieved on four consecutive trials, a RED light located in the CTT steering column display is activated, and the driver must wait 10 minutes before attempting the test again. If the car is driven during this period the alarms will operate.

C. ANTICIRCUMVENTION

Various measures have been incorporated into the DDWS to prevent cheating. These include seals on the equipment, a seat weight sensor, and the logic that is programmed into the DDWS computer. The hazard lights or flasher, the horn, battery and alternator connections, and the fuse box are sealed with a molten rubber material that is imprinted with a Department of Transportation stamp before it hardens. The cables to the CTT electronics module located in the trunk are threaded with wires and sealed with a lead seal that has a DOT imprint. The screws in the threshold plates that cover the wires going from the computer electronics module in the trunk to the CTT located on the vehicle steering column are coated with red nail polish. All of these seals can be easily removed, but it would be immediately obvious.

The driver seat is equipped with a seat weight sensor consisting of a strain gauge installed in the springs of the seat. This measures weight and weight distribution. At the time of each trial the data logger records whether or not the operator's weight distribution falls within a pre-determined range.

The logic of the system alone can prevent many forms of cheating. For instance, it would be virtually impossible for anyone other than the intended driver to drive the car in a normal mode. Previous research has shown that it takes approximately 300 trials on the CTT to reach one's maximum skill level. It would take hours to train an individual who had to wait 10 minutes every time he/she failed the test. It would also be apparent what was happening by a simple examination of the data tape.

Also, the system is designed to prevent the operator from driving in an impaired state, while at the same time minimizing inconvenience to the sober driver. For example, the system requires a new test every time the driver leaves the driver seat. This prevents the trained subject from passing the test and then switching with a friend. However, if the test has been passed, and the car driven, the driver can turn off the engine and remain in the seat for 10 minutes without retesting. This allows the driver to get gas at a full service station without having to retake the test.

Once the driver presses the test button and starts a CTT test sequence any one of four conditions can lead to a 10 minute wait. The four conditions are: 1) exit the car, 2) start driving, 3) turn the ignition off, or 4) fail the test strategy. Once the test series is started the test must be passed or else there will be a 10 minute fail period. The driver cannot partially complete the CTT test strategy, turn off the ignition, turn it back on and have four more tries. If the driver must leave the car for any reason, the test sequence must be completed if initiated. Upon returning, he/she will have to take the test again, but there will not be a 10 minute waiting period. Should the driver forget to take the test and start driving, the alarms will operate; however, this does not bring on the waiting period. Presumably, the driver will hear the horn and be reminded to stop and take the test.

SECTION VII

EPILOGUE

The purpose of this Volume I of the final report (Contract DOT-HS-8-02052, "Field Test of the Drunk Driving Warning System") was to summarize research on in-vehicle alcohol safety devices leading up to the Drunk Driving Warning System (DDWS) concept, and to describe the current DDWS configuration being tested in California.

Volume II of the report describes the California field test of the DDWS concept in a judicial setting with convicted drunk drivers. The purpose of the field test was to investigate the feasibility of the DDWS concept for its practicality, public acceptability and effectiveness. Practicality issues investigated involved legislative, court and probation requirements, cost, implementation procedures, etc. Public acceptability issues investigated involved user reactions to the concept including those of court officials, police, probation personnel, defense counsel, convicted drunk drivers, family, friends, etc. Effectiveness issues investigated involved ability of the DDWS to influence drinking behavior in a positive manner and/or to deter drunk-driving trips.

The combined Volume I and II report is intended to provide NHTSA with quantitative information on the credibility of the DDWS concept, its reliability, procedures for its implementation in a judicial setting and its ability to deter drunk-driving trips. Based on data emerging from this work, NHTSA hopes to make recommendations to government agencies and private industry for use of in-vehicle alcohol safety devices such as the DDWS, among known drinking drivers and/or the general public.

REFERENCES

Allen, R. Wade, and Henry R. Jex, Feb. 1980. "Driving Simulation — Requirements, Mechanization and Application, SAE Paper 800448, presented at Congress and Exposition, Detroit.

Allen, R. Wade, and Henry R. Jex, May 1973. Visual Motor Response of Crewmen During a Simulation 90-Day Space Mission as Measured by the Critical Task Battery, NASA CR-2240.

Allen, R. Wade, Raymond E. Magdaleno, and Henry R. Jex, Oct. 1975. Effects of Wideband Auditory Noise on Manual Control Performance and Dynamic Response, AMRL-TR-75-65.

Allen, R. W., S. H. Schwartz, A. C. Stein, et al., Apr. 1978. The Effects of Alcohol and Marihuana on Driver Control Behavior. Vol. I: Laboratory Simulation Experiment, Systems Technology, Inc., ITR-1066-1. (See footnote)*

Hovland, Carl I., 1951. "Human Learning and Retention," in S. S. Stevens, ed., Handbook of Experimental Psychology, Wiley, New York, pp. 613-689.

Hurst, P. M., 1973. "Epidemiological Aspects of Alcohol in Driver Crashes and Citations," J. Safety Research, pp. 130-148.

Jex, Henry R., Richard J. DiMarco, and Warren F. Clement, May 1977. Effects of Simulated Surface Effect Ship Motions on Crew Habitability — Phase II. Vol. 3: Visual-Motor Tasks and Subjective Evaluations, Naval Sea Systems Command PMS-304.

Jex, H. R., J. D. McDonnell, and A. V. Phatak, Nov. 1966. A "Critical" Tracking Task for Man-Machine Research Related and the Operator's Effective Delay Time. Part I: Theory and Experiments With a First-Order Divergent Controlled Element, NASA CR-616.

Jex, H. R., and R. A. Peters, 27 Aug. 1974. Technical Specifications for a Critical Task Interlock, (Revision of Attachment A to DOT/Transportation Systems Center Contract No. DOT-TSC-848 Alcohol Safety Interlock System), Systems Technology, Inc., WP-1054-1.

McDonnell, J. D., and H. R. Jex, Jan. 1967. A "Critical" Tracking Task for Man-Machine Research Related to the Operator's Effective Delay Time. Part II: Experimental Effects of System Input Spectra, Control Stick Stiffness, and Controlled Element Order, NASA CR-674.

---

*CTT data unpublished, however, reference describes experiment during which CTT data was collected.

Moskowitz, H., and D. Depry, 1968. "Differential Effect of Alcohol on Auditory Vigilence and Divided Attention Tasks," Quarterly J. of Studies on Alcohol, Vol. 29, pp. 54-63.

Oates, John F., Jr., 1973. Experimental Evaluation of Second-Generation Alcohol Safety-Interlock Systems, Transportation Systems Center DOT-TSC-NHTSA-73-9.

Oates, John F. Jr., June 1975. David F. Preusser, Richard D. Blomberg, and Marlene S. Orban, June 1975a. Laboratory Testing of Alcohol Safety Interlock Systems. Vol. I: Procedures and Preliminary Analyses, Dunlap and Associates, Inc.

Oates, John F. Jr., David F. Preusser, and Richard D. Blomberg, Dec. 1975b. Laboratory Testing of Alcohol Safety Interlock Systems, Phase II, Dunlap and Associates, Inc.

Parzen, E., Modern Probability and Its Applications, Wiley, New York, 1960.

Peters, Richard A., Henry R. Jex, and Ronald J. Fifer, Nov. 1975. Operations Manual: Critical Task Interlock, Systems Technology, Inc., TM-1054-1.

Reynold, George S., A Primer of Operant Conditioning, Glenview, Illinois, Scott, Foresman, 1968.

Sussman, Donald, and Charles N. Abernethy III, Apr. 1973. Laboratory Evaluation of Alcohol Safety Interlock Systems. Vol. I: Summary Report, Transportation Systems Center, DOT-TSC-NHTSA-73-3.

Tennant, Jean A., and Richard R. Thompson, Jan. 1973. "A Critical Tracking Task as an Alcohol Interlock System," SAE Paper 730095, presented at Int'l. Automotive Engineering Congress and Expo., Detroit.

Transportation Systems Center, 1973. Evaluation of a Vehicle Mounted Breath Testing Device, In-House Report.

Voas, R. B., May 1974. Alcohol, Drugs and Young Drivers, National Highway Traffic Safety Administration, Traffic Safety Programs.

Wald, A., Sequential Analysis, Wiley, New York, 1962.

Woodworth, R. S. and H. Schlosberg, 1964. Experimental Psychology, Holt, Rinehart & Winston, New York.

APPENDIX A

CTT DATA ANALYSIS

OVERVIEW

Since the CTT was developed in the mid 1960s, several data bases have been generated that can tell us something about the statistical properties of CTT scores and their sensitivity to alcohol impairment. This appendix will review past CTT data base analyses, and present results from the reanalysis of some recent CTT experiments.

In order to establish a framework for data analysis, we first hypothesize a statistical model for CTT scores which includes both deterministic and random components. This model then provides a structure for considering results of Analysis of Variance procedures (ANOV) applied to past CTT data bases. Next, we consider detailed reanalysis of recent CTT data bases using ANOV and regression analysis to establish the dependency of CTT score on BAC (blood alcohol concentration) and the variability of this relationship between subjects. Finally, a simple statistical model is set forth that is used in Appendices B and C to analyze the effect of various decision strategy parameters.

GENERAL STATISTICAL MODEL

In order to accomplish the above we will review and reanalyze a variety of past CTT data bases using a general statistical model that accounts for variety of influences on CTT scores. A general model is considered here that partitions the scores into a series of deterministic and random components:

$$\underbrace{\lambda_c}_{\text{Net Score}} = \underbrace{\lambda_{c_o}}_{\text{Population Average}} + \underbrace{\Delta\lambda_1}_{\text{Stress Impairment}} + \underbrace{\Delta\lambda_2}_{\text{Trial Order}} + \underbrace{\Delta\lambda_3}_{\text{Long Term Trends}} \Bigg\} \text{ Deterministic Effects}$$

$$+ \underbrace{\varepsilon_1}_{\text{Within a Subject}} + \underbrace{\varepsilon_2}_{\text{Between Subjects}} + \underbrace{\varepsilon_3}_{\text{Subjects by Stress Interaction}} \Bigg\} \text{ Random Effects} \tag{A-1}$$

The deterministic effects describe what happens to the performance metric on the average. Here we have allowed for various effects:

- Stress applied to the human operator which would typically degrade performance.
- Trial order or short-term trends having to do with warmup and/or fatigue effects causing a trend across a few repeated trials.
- Long-term trends due to learning with repeated task experience.

Random effects concern statistical variations in performance, without any apparent causal factors:

- Within a subject - random variations across multiple trials which are over and above general trial-to-trial trends.
- Between subjects - random average performance differences between randomly selected subjects.
- Subject by stress interaction - random differences in response to a given stress between randomly selected subjects.

Equation A-1 can be considered an Analysis of Variance (ANOV) model, and estimates of the various model components can be obtained using ANOV and regression analysis procedures on a given data base.

MODEL COMPONENT ANALYSIS

The Critical Tracking Task (CTT) has been used in a variety of research studies for more than a decade and good estimates of various Eq. A-1 terms are available. In Fig. A-1, for example, we see the effect of alcohol on CTT score decrements compared across several experiments conducted by different research groups with different experimental procedures. Note that CTT performance universally degrades by about 10 percent at the common legal limit of blood alcohol concentration (BAC) specified in many state vehicle codes (i.e., 0.10 percent), and more than twice this decrement at BACs above 0.20 percent, characteristic of heavy drinking.

Short- and long-term trends in CTT scores have been analyzed in several experiments. Allen and Jex (1973) analyzed the psychomotor performance of four subjects who were isolated in a space station simulator for three months. CTT scores were obtained two times a day for 12 weeks. ANOV procedures did not show any day-to-day trends averaged across weeks, but did show a week-to-week improvement trend over the twelve weeks. We have also reanalyzed some other more recent CTT data bases (i.e., Oates, 1973; Oates, et al., 1975a, 1975b). In this reanalysis we allowed for a trial-to-trial component for six trial blocks. The analysis showed minimal trial-to-trial effects.

The random variance components in Eq. A-1 can be estimated with ANOV procedures. This has been done for a wide variety of CTT experiments, and the results are summarized in Table A-1. Due to equipment and training variations, the mean performance level varied considerably between some experiments. To minimize this factor normalized variability components are also given in Table A-1, and it is the normalized levels that (given as percentage) are noted to be quite consistent across various experiments. One standard deviation of within-subject variability runs about 9 percent (range of 6-10 percent), with between-subject variability running about the same level. Subject-by-stress interaction runs somewhat less at about 5 percent on the average (range of 3-6 percent).

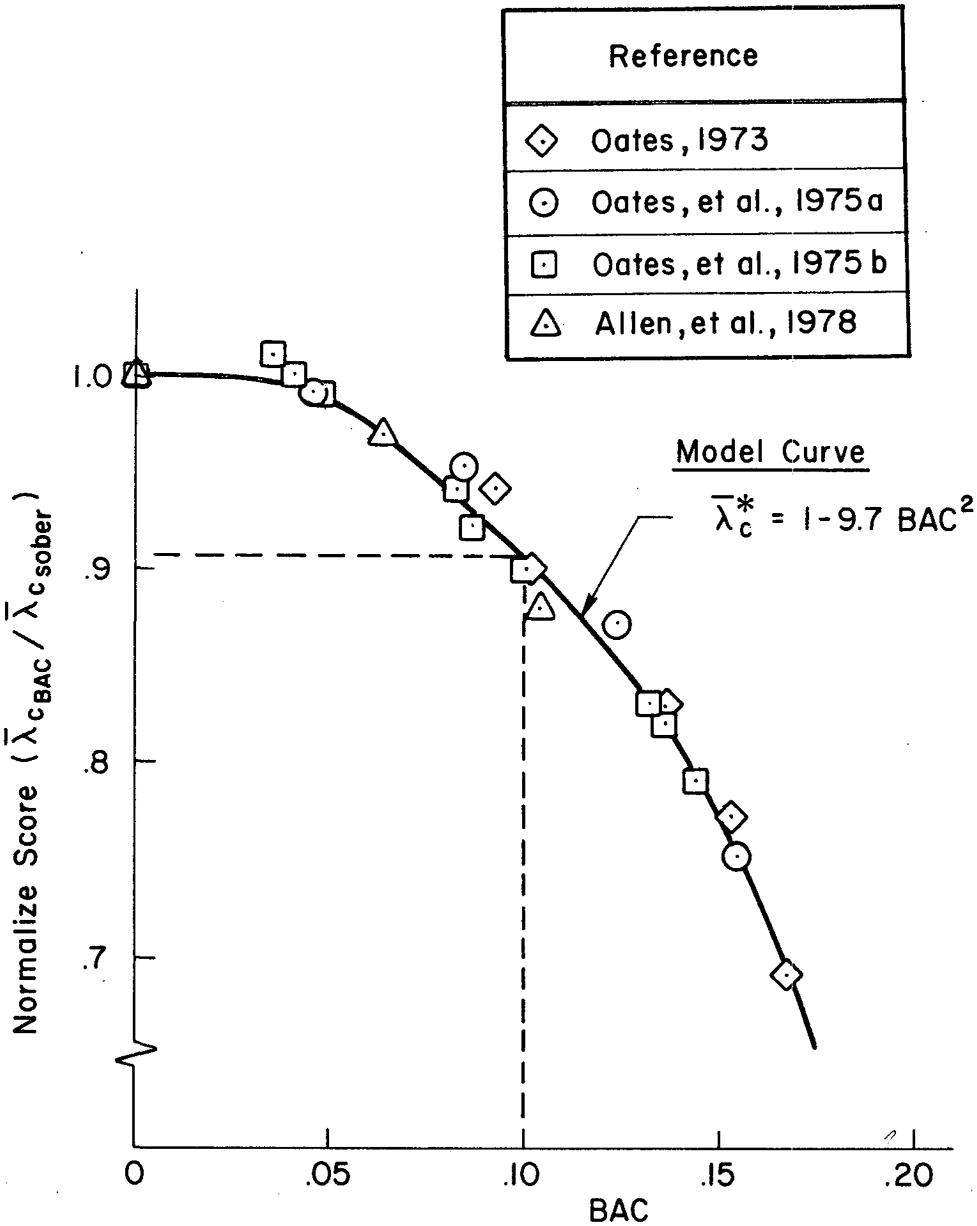


Figure A-1. Critical Task Score as a Function of BAC Across Several Experiments. Data Normalized to Account for Differences in Task Mechanization, Procedures, etc.

TABLE A-1. ANALYSIS OF VARIANCE FOR CTT DATA

REFERENCE	NO. OF Ss	$\bar{\lambda}$	ABSOLUTE VARIABILITY, σ			NORMALIZED VARIABILITY, $\sigma/\bar{\lambda}_c$		
			WITHIN Ss σ_{ε_1}	BETWEEN Ss σ_{ε_2}	SUBJECT × STRESS σ_{ε_3}	WITHIN Ss $\sigma_{\varepsilon_1}/\bar{\lambda}_c$	BETWEEN Ss $\sigma_{\varepsilon_2}/\bar{\lambda}_c$	SUBJECT × STRESS $\sigma_{\varepsilon_3}/\bar{\lambda}_c$
ALCOHOL STRESS								
Oates, 1973, ASIS2								
a) Analyzed as 2 groups (female vs. male)	14	4.75	.436	.396	.297	9.2%	8.3%	6.3%
b) Analyzed as 3 groups (random)	15	4.75	.421	.437	.293	8.9%	9.2%	6.2%
c) Analyzed as 5 groups (age effects)	15	4.75	.434	.418	.288	9.1%	8.8%	6.1%
Oates, et al., 1975a	24	4.64	.434	.476	.263	9.4%	10.3%	5.7%
Oates, et al., 1975b	24	4.62	.442	.431	.197	9.6%	9.3%	4.3%
Allen, et al., 1978	12	4.7	.32	.33	.28	6.8%	7.0%	6.0%
OTHER STRESSES								
Allen, et al., 1975 Broadband Noise	9	3.1	.28	.18	.12	9.1%	5.8%	3.9%
Allen and Jex, 1973 90 Day Confinement	4	6.5	.41	.319	.333	6.2%	4.9%	5.1%
Jex, et al., 1977 Motion Effects	8	5.14	.514	.717	.134	10.0%	14.0%	2.7%

The Table A-1 results tell us several things about setting up a decision-making strategy using the Critical Tracking Test. First, within-subject variability is about the same magnitude as the degradation due to legally drunk blood alcohol levels. Thus, detecting this magnitude of change in performance level will definitely require multiple samples for reasonable reliability. Second, the between-subject variability component represents a significant portion of the total variability so that each subject should have an individualed performance criterion. Finally, after accounting for within- and between-subject variability components, a small amount of residual variability between subjects is still to be expected in their differential responses to a given stress.

In order to develop a statistical model for Critical Tracking Task scores, the Dunlap data bases summarized in Table A-1 (Oates, 1973; Oates, et al., 1975a, 1975b) were subjected to a detailed reanalysis. Individual Score decrements were obtained by removing each subject's mean sober score from all his/her individual scores. The ensemble distributions of the resulting differential scores at four BAC levels for two experiments are shown in Fig. A-2. The cumulative Gaussian distribution scaling on Fig. A-2 (standard "probit" plots) helps reveal several characteristics of the CTT score distributions. First, the scores at each BAC are roughly Gaussianly distributed between the 5 percent and 95 percent points, as indicated by the straight slopes. Second, the variability of population decrements increases slightly with BAC (shallower slope). Since we have accounted for the between-subject variability by analyzing differential scores, the increasing variability with BAC amounts to a combination of increased trial-to-trial variability (ε_1) combined with subject-by-stress interaction variability (ε_3) discussed previously.

REANALYSIS OF DUNLAP DATA

Under contract to the Transportation Systems Center, Dunlap and Associates conducted three laboratory evaluations of the Critical Tracking Task (Oates, 1973; Oates, et al., 1975a, 1975b). The experiments

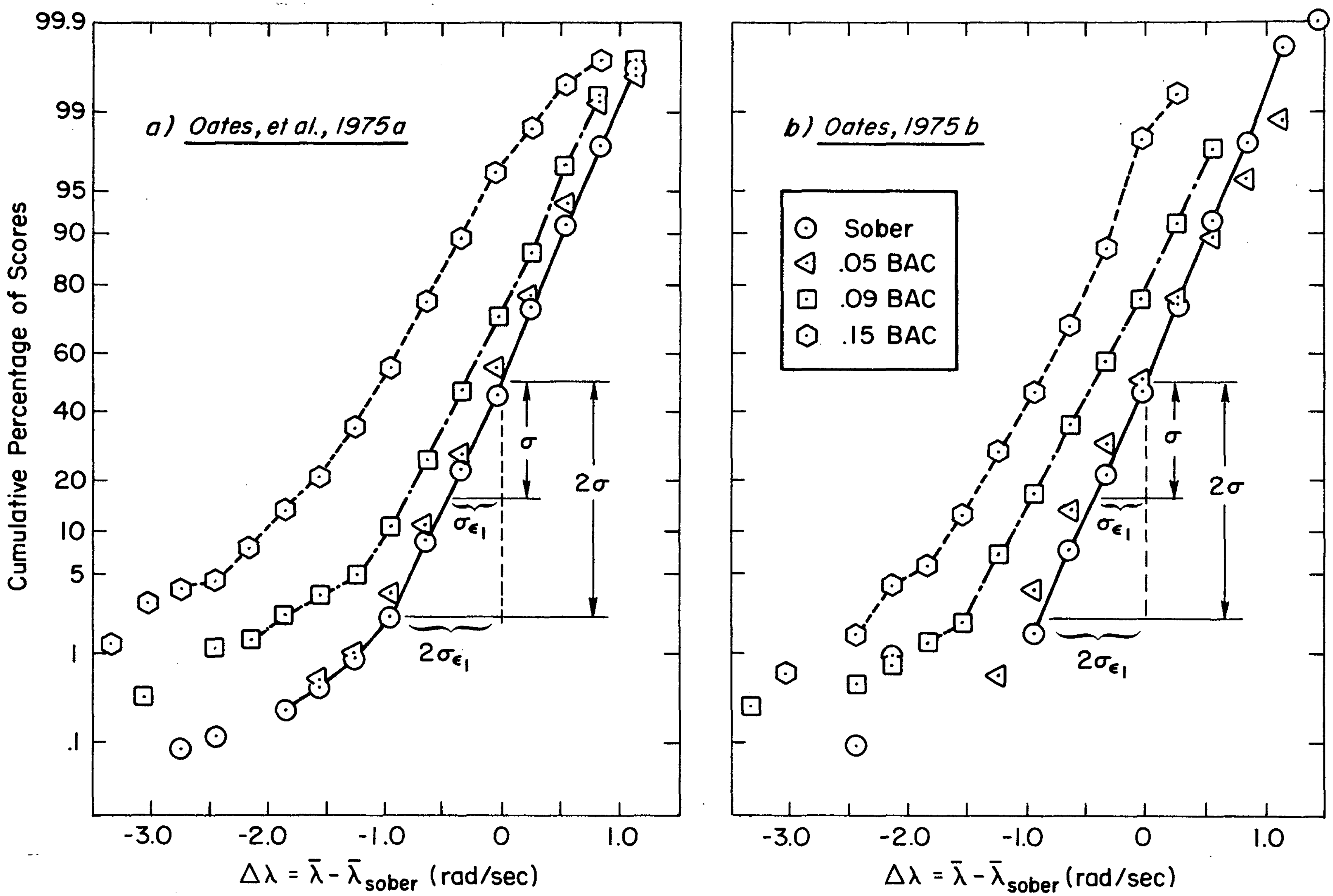


Figure A-2. Cumulative Distribution Functions of Normalized λ_c Scores from the Dunlap Experiments were conducted in sessions which occurred every other day. Subjects performed successive repititions of the critical task in blocks. Six or seven blocks were performed within a session. All sessions were given in the late afternoon or evening hours.

In the first two experiments the sessions were divided between alcohol treatment sessions and placebo or control sessions. During the alcohol treatment sessions alcohol consumption was interspersed between blocks so that a subject's BAC level would increase with each block of trials and the peak BAC level was reached on the last block of trials. In the third experiment the sessions were divided into four treatment conditions representing peak alcohol levels of 0, 0.05, 0.10, and 0.15 percent BAC. In this experiment the blocks occurred prior to drinking at peak BAC level, 30 minutes after the peak level, and 60 minutes after the peak level.

Mixed sexes were used in the first experiment and only males were used in the latter two experiments. All subjects were heavy drinkers. Training consisted of 96 to 100 trials given in two or three sessions. Incentives were used throughout the training and the experimental sessions.

Repeated measures analysis of variance (ANOV) tests were performed on the alcohol and control treatment sessions separately. This eliminated the confounding of BAC level with the block order effects. BAC level was the only significant effect (at the 0.001 level) except for a session order effect in the first experiment. The block order and trial order were not significant in any of the experiments. Breakdown of the orthogonal components for the BAC level showed that the linear component accounts for 90 to 95 percent of the variance.

The data in Fig. A-1 show that alcohol effects on CTT score are closely approximated by a $(BAC)^2$ function. Furthermore, for a given BAC, the Fig. A-2 results show that CTT scores are approximately normally distributed. Based on these results, a regression analysis was performed on the three Dunlap data bases using $(BAC)^2$ as the independent variable. Regression coefficients were obtained on a subject-by-subject basis and the results are summarized in Table A-2. The correlation coefficients indicate a fairly good agreement between CTT performance scores and the regression function

$$\bar{\lambda}_c = a + b(\mathrm{BAC})^2$$

The within-subject variability (S_w) is consistent with the ANOV results in Table A-1 (i.e., $\sigma_{\varepsilon_1} \cong 0.43$ for the three Dunlap experiments). The between-subject variability in the 0 BAC scores ($\bar{\lambda}_c$) and regression intercepts (a) is also consistent with the Table A-1 between-subject variabilities (σ_{λ_2}). Finally, the between-subject variability in the regression slope, b, gives us a measure of the subject-by-stress interaction, σ_{ε_3}, which is a function of BAC^2. The results will be useful for quantifying a CTT performance score statistical model as discussed below.

ANALYTICAL MODEL

The preceding data analysis has shown some consistent trends which can be expressed analytically as a function of BAC. This statistical model will then be useful in Appendices B and C for analyzing the effect of various test strategies on the ability of the CTT to discriminate between sober and drunk drivers. In Fig. A-1 a normalized regression function was shown to give good agreement between CTT score decrement and BAC:

$$\lambda_c^* = \frac{\lambda_{c\ \mathrm{BAC}}}{\lambda_{s\ \mathrm{sober}}} = 1 - 9.7\ \mathrm{BAC}^2 \qquad \text{(A-2)}$$

The mean sober λ_c score across all three Dunlap experiments (Table A-2) is $\bar{\lambda}_c = 4.92$, which results in the regression equation

$$\bar{\lambda}_c = 4.92 - 47.7\ \mathrm{BAC}^2 \qquad \text{(A-3)}$$

The BAC^2 coefficient above is in close agreement with the average slope term determined from the subject-by-subject regression of Table A-2 ($\bar{b} = 48.2$).

TABLE A-2. SUBJECT-BY-SUBJECT REGRESSION ANALYSIS OF DUNLAP DATA

SUBJECT	OATES, 1973						OATES, et al., 1975a						OATES, et al., 1975b					
	0 BAC		BAC^2 REGRESSION				0 BAC		BAC^2 REGRESSION				0 BAC		BAC^2 REGRESSION			
	$\bar{\lambda}_c$	S_w	a	b	S_r	r	$\bar{\lambda}_c$	S_w	a	b	S_r	r	$\bar{\lambda}_c$	S_w	a	b	S_r	r
1	5.67	0.41	5.67	-59.9	0.415	0.819	5.29	0.52	5.22	-106.4	0.697	0.760	4.55	0.525	4.57	-39.9	0.466	0.506
2	4.81	0.38	4.83	-53.0	0.501	0.713	4.36	0.40	4.39	-49.4	0.462	0.668	5.98	0.564	6.07	-82.9	0.528	0.742
3	4.97	0.50	4.94	-65.3	0.503	0.746	4.41	0.31	4.43	-107.7	0.460	0.636	5.17	0.524	5.26	-78.8	0.507	0.714
4	5.54	0.37	5.54	-79.4	0.362	0.916	4.63	0.40	4.60	-71.7	0.398	0.718	4.84	0.401	4.80	-49.6	0.428	0.671
5	4.88	0.52	4.90	-47.1	0.467	0.690	5.26	0.63	5.29	-56.2	0.605	0.607	4.99	0.419	4.83	-48.3	0.477	0.640
6	4.83	0.35	4.51	-28.7	0.416	0.587	4.82	0.39	4.88	-50.7	0.433	0.709	5.47	0.377	5.53	-53.8	0.360	0.816
7	4.60	0.34	4.59	-42.6	0.372	0.769	5.15	0.46	5.18	-35.1	0.480	0.515	4.31	0.618	4.34	-54.2	0.581	0.693
8	5.01	0.37	5.01	-26.5	0.373	0.584	5.58	0.49	5.62	-43.8	0.530	0.684	4.53	0.388	4.59	-50.9	0.465	0.600
9	5.34	0.34	5.32	-58.1	0.353	0.771	5.16	0.46	5.17	-48.1	0.512	0.678	5.36	0.537	5.28	-28.6	0.641	0.238
10	5.85	0.38	5.87	-43.6	0.394	0.825	5.11	0.44	5.08	-30.9	0.444	0.546	5.06	0.306	5.04	-27.6	0.382	0.395
11	4.76	0.53	4.72	-62.0	0.557	0.567	5.71	0.43	5.75	-58.3	0.421	0.723	5.47	0.434	5.41	-50.4	0.560	0.487
12	4.30	0.43	4.33	-49.8	0.501	0.758	4.23	0.33	4.22	-32.0	0.369	0.559	4.63	0.476	4.61	-29.0	0.518	0.411
13	5.27	0.40	5.34	-34.4	0.486	0.595	4.82	0.60	4.85	-50.3	0.582	0.548	4.14	0.362	4.40	-27.0	0.386	0.514
14	4.73	0.40	4.73	-53.4	0.500	0.776	4.90	0.44	4.91	-47.8	0.451	0.650	4.48	0.432	4.43	-34.1	0.430	0.506
15	5.00	0.40	5.00	-43.4	0.390	0.773	4.57	0.39	4.55	-36.4	0.406	0.398	4.53	0.428	4.43	-76.8	0.453	0.741
16	5.03	0.41	5.01	-53.4	0.440	0.829	4.83	0.52	4.84	-37.3	0.480	0.476	5.00	0.388	4.97	-29.6	0.403	0.494
17							5.10	0.44	5.13	-41.3	0.403	0.693	4.91	0.383	4.88	-19.6	0.377	0.371
18							3.78	0.44	3.79	-29.1	0.396	0.577	4.65	0.403	4.59	-57.7	0.448	0.676
19							4.24	0.44	4.34	-24.5	0.498	0.482	5.40	0.413	5.34	-55.7	0.410	0.777
20							4.59	0.47	4.60	-46.7	0.526	0.679	4.18	0.508	4.27	-54.1	0.509	0.549
21							5.20	0.40	5.23	-27.6	0.401	0.594	5.06	0.410	5.05	-31.0	0.431	0.516
22							4.71	0.37	4.73	-39.4	0.360	0.739	4.96	0.409	4.95	-57.2	0.413	0.595
23							5.76	0.47	5.83	-56.3	0.485	0.701	4.82	0.310	4.84	-32.7	0.369	0.588
24							4.48	0.48	4.44	-37.6	0.497	0.545	4.43	0.493	4.76	-44.4	0.515	0.580
Mean	5.04	0.41	5.02	-50.4	0.439		4.86	0.45	4.88	-48.5	0.471		4.87	0.438	4.89	-46.4	0.461	
S_B	0.41		.43	13.7			0.49		0.50	22.9			0.46		0.44	17.2		

Regression Function: $\lambda_c = a + b\,(BAC)^2$

S_w = within-subject sample standard deviation

S_B = between-subject sample standard deviations

S_r = sample standard deviation about regression line

CTT scores have three variability components as indicated in Eq. A-1. The sum of the variances of these components gives the total variance of CTT scores

$$\sigma_\lambda^2 = \sigma^2_{\varepsilon_1} + \sigma^2_{\varepsilon_2} + \sigma^2_{\varepsilon_3} \tag{A-4}$$

Variances for within ($\sigma^2_{\varepsilon_1}$) and between ($\sigma^2_{\varepsilon_2}$) subject variability are approximately equal as averaged across the Dunlap data bases:

$$\begin{aligned} \sigma^2_{\varepsilon_1} &= \sigma^2_{\varepsilon_2} = 0.212 \\ \sigma_{\varepsilon_1} &= \sigma_{\varepsilon_2} = 0.46 \end{aligned} \tag{A-5}$$

The between-subject variability of the regression slopes gives an expression for subject-by-stress variability (σ_{ε_3}) that is dependent on BAC^2. Across all three Dunlap experiments, if we ignore 2 outliers in the second study, we find that

$$\sigma_{\varepsilon_3} = 15(BAC)^2 \tag{A-6}$$

As a final check on equations (A-3 through A-6), distributions for the Dunlap data base were obtained after removal of each subject's mean sober performance level (i.e., this effectively made $\bar{\lambda}_c = 0$ and $\sigma_{\varepsilon_2} = 0$ in Eqs. A-3 and A-4). The differential mean and total standard deviations for each experiment are plotted in Fig. A-3 along with the model curves for $\Delta\lambda_c$ and σ_λ. The differential mean data show good agreement with the model. The standard deviation data do not show as close model agreement as the mean data, but are still consistent in their BAC dependency.

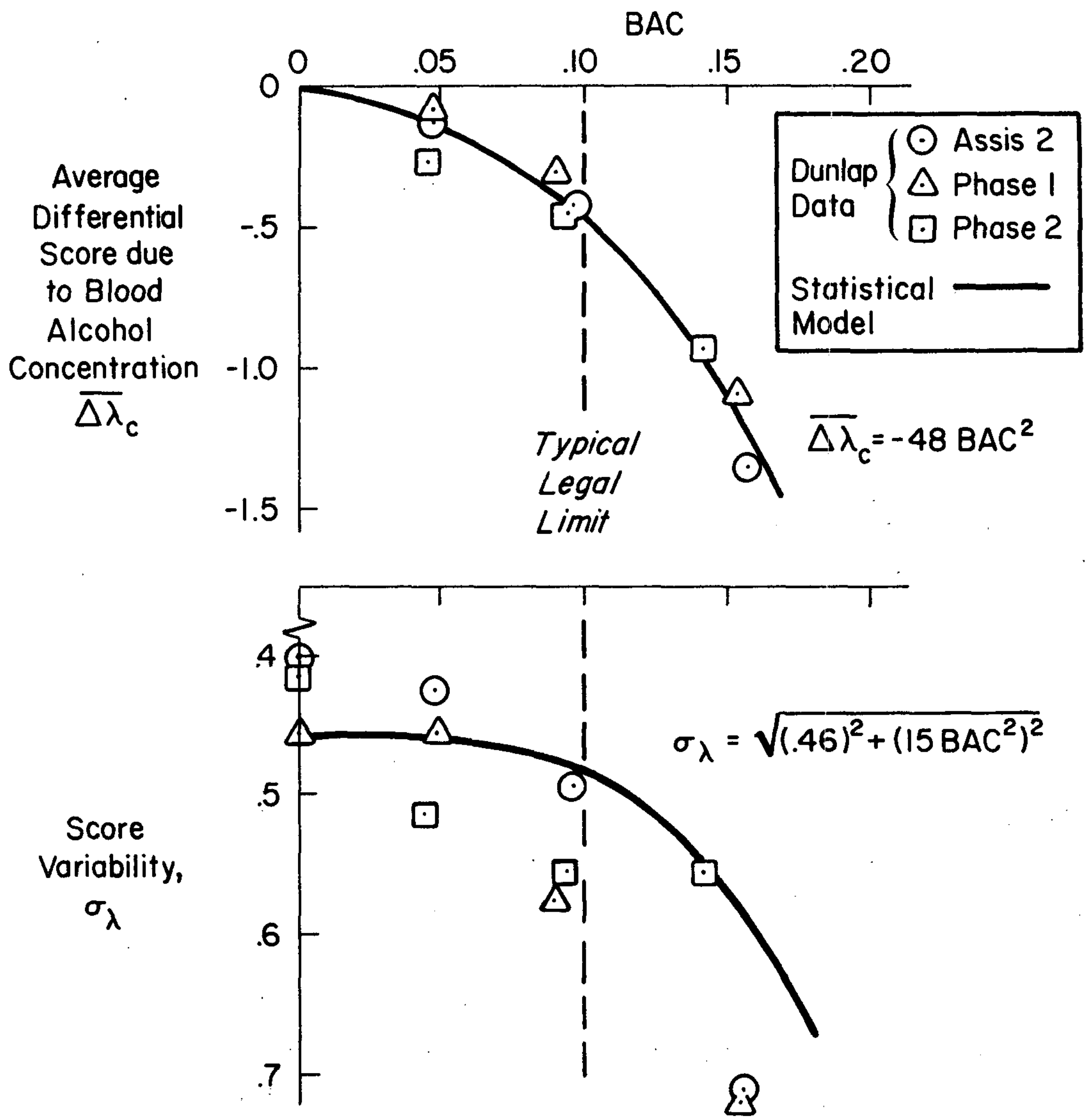


Figure A-3. BAC Effect on the Differential Mean and Variability of CTT Scores

APPENDIX B

TEST DISCRIMINABILITY ANALYSIS

OVERVIEW

Given a fair amount of knowledge about the statistical properties of CTT scores as discussed in Appendix A, we now would like to set up a decision strategy that will allow us to detect alcohol-impaired psychomotor behavior. Given that an individual's trained, sober performance capability is known, we wish to be able to detect when performance is reliably below this criterion level. This presents a signal-in-noise detection problem; i.e., to detect the mean shift in score in the face of its intrinsic variability. This will require a multiple-trial strategy since, as mentioned previously, the mean shift in score due to alcohol impairment to be detected is roughly the same magnitude as the standard deviation in CTT scores.

In the detection problem one wishes to minimize two types of classical decision-making errors. The first, referred to as a "Type I" error (usually designated by the symbol α), is the probability that there is no difference in the mean performance (i.e., rejecting the null hypothesis) when the test outcome says there is a decrement. In the current application this would amount to a sober and unimpaired subject being rejected by the test sequence. The second, or "Type II" error (β), is the probability that there is a decrement (hence, impairment) when the test concludes there is none. Here, β, would be the probability that an alochol- or drug-impaired subject could pass the test sequence.

The approach taken in setting error probabilities is similar to the general experimental design problem, in that we will assign a tolerable level for α, setting it at a relatively low value so that the unimpaired operator is not inconvenienced (i.e., can routinely pass the test). However, because test time is costly, one cannot simply take an arbitrary number of samples in order to achieve a low β level for reliably detecting a specified change in performance $\Delta\lambda$. In order to have a practical psychomotor screening test, one has to pick an <u>efficient</u>

multiple trial strategy and attempt to achieve a low level of β at reasonable impairment levels with a relatively few number of trials on the average, all subject to an acceptably low α. This approach is discussed further below.

DISCRIMINABILITY AND PASS LEVEL

The basic discriminability problem can best be illustrated by the test score cumulative distribution shown in Fig. B-1 based on the Appendix A statistical model. A pass level has been selected to give a low sober failure rate (α = 2.5 percent). Given this pass level, on any given trial the average subject has a 17 percent chance of failing the test at 0.10 BAC and a 56 percent chance of failing the test at 0.15 BAC. These failure levels are obviously not very high in terms of effectively discriminating against drunk drivers. This state of affairs can be improved with multiple-trial test strategies, however, as discussed below.

A key point to note here is that multiple trial test strategies will change the sober failure rate (α) for a given pass level. In optimizing the test strategy we wish to hold α constant, however, at a given specified level (similar to the classical experimental design problem). Thus the test score pass level must be varied in a specific way with each strategy in order to maintain a constant α. This point is critical to optimizing test discriminability, and appears to have been overlooked by previous investigators.

In general, test score pass levels, λ_p, are set some multiple K_σ of the test score standard deviation below the individual's sober baseline, so as to achieve a specified α based on Gaussian statistical assumptions (or from probit plots such as those in Appendix A, Fig. A-2):

$$K_{\sigma_n} = \frac{\bar{\lambda}_{sober} - \lambda_{pass}}{\sigma_{\lambda_n}} \qquad \text{(B-1)}$$

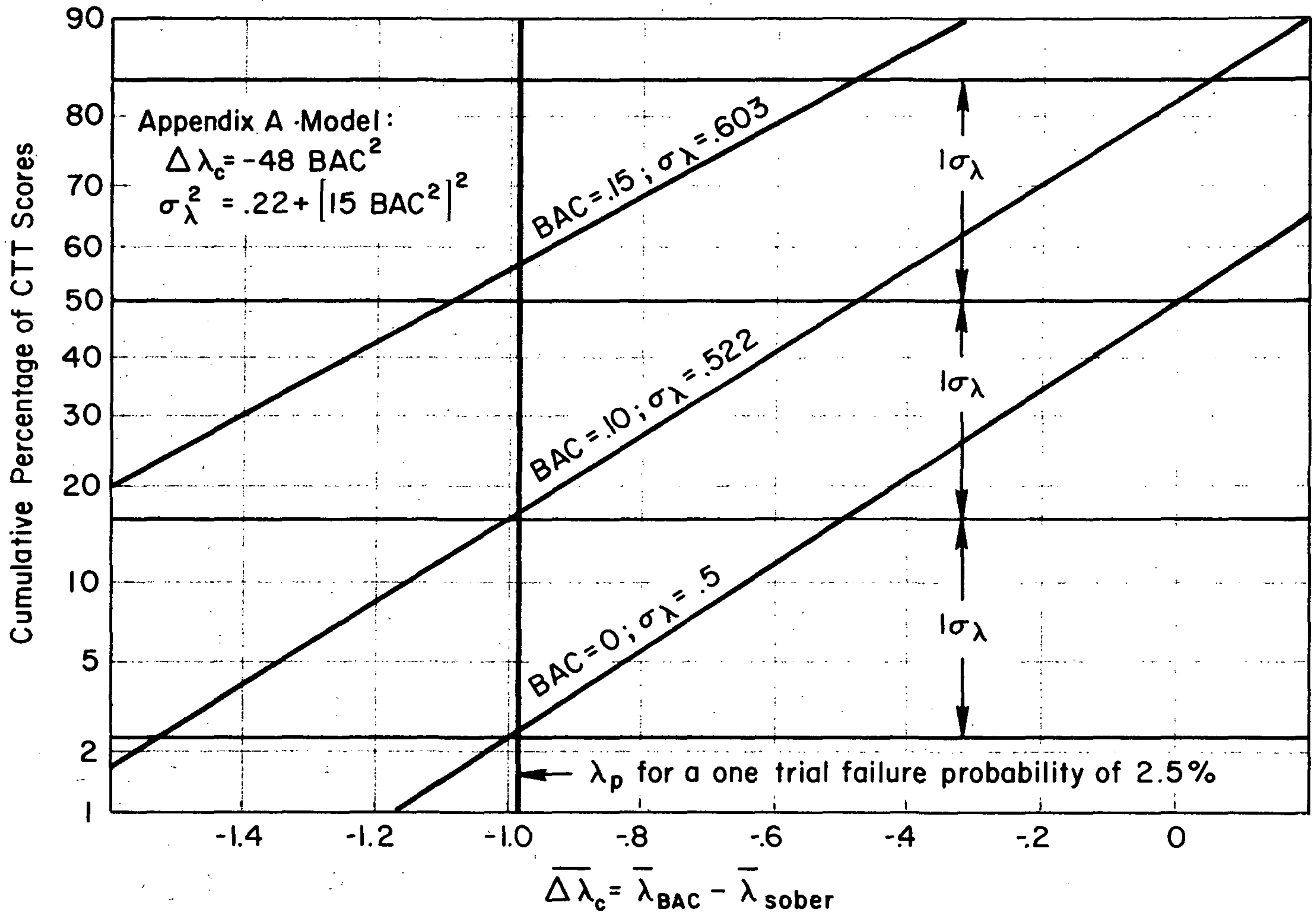


Figure B-1. CTT Gaussian Score Distributions According to Appendix A Model

This can be solved for the pass criterion:

$$\lambda_p = \bar{\lambda}_{sober} - K_{\sigma_n}\sigma_\lambda \tag{B-2}$$

where

λ_p = criterion level of λ_c above which a trial is passed

$\bar{\lambda}_{sober}$ = a given subjects trained, unimpaired performance level

σ_λ = standard deviation of scores for the test strategy used

K_{σ_n} = constant depending on the test strategy and number of trials, n; and desired level of α

For example, looking back at Figure B-1, for $\alpha \doteq 0.025$ (sober fail probability at 2.5 percent) the <u>single</u> trial score decrement for sober tests (BAC = 0) is about ($\Delta\lambda \doteq$)$2\sigma_\lambda$ below the sober mean; hence, $K_{\sigma_1} \doteq 2.0$. Thus the score decrement which could be reliably detected in a single trial (α = 0.025) is $\Delta\lambda = -2\sigma_\lambda \doteq -0.98$ rad/sec. This is much larger than the desired BAC = 0.10 decrement of about -0.48, so multiple trials will be required to increase test sensitivity. The "probability laws" for various multiple-trial test strategies are given below.

TEST STRATEGY PROBABILITY LAWS

The sober failure rate or Type I error (α) is set according to the distribution function for CTT scores and the probability law for a given test strategy. The probability laws can be derived simply by assuming that the test scores for each trial are independent random variables with equal distribution functions. Given these assumptions the probability laws for the test strategy categories discussed in the main text can be derived as follows.

Average of n trials

The sampling distribution for the average of n samples is a well known statistical result (Parzen, 1960). The mean of the averages is equal to the mean of the original distribution, and the variance is reduced by a factor of 1/n over the single trial distribution:

$$\bar{\lambda}_n = \bar{\lambda} \tag{B-3}$$

and

$$\sigma^2_{\lambda_n} = \sigma^2_{\lambda}/n \tag{B-4}$$

Thus for a cumulative Gaussian distribution plotted on probability paper as in Fig. B-1, the distribution for the average of n samples will be rotated about the mean or 50 percent point.

Since the standard deviation of the mean is reduced over the single trial value, the pass criterion coefficient in Eq. E-2 is also similarly reduced

$$K_{\sigma_n} = K_{\sigma_1}/\sqrt{n} \tag{B-5}$$

where K_{σ_n} is the value required to achieve a given α for the average of n trials strategy.

The average-of-n trial strategy was applied to the combined latter two Dunlap data bases (Oates, et al., 1975a, 1975b) using each subject's sober levels of $\bar{\lambda}_c$ and σ_λ. The results are listed in Table B-1 and plotted in Fig. B-2. The features of these results are as follows. First the achieved sober failure rates varied from 3-5 percent (somewhat higher than the desired 2.5 percent). This is probably due to the low score tail of the λ_c distribution (Appendix A, Fig. A-2), which indicates a higher occurrence of scores than would be predicted by a normal distribution. The Fig. B-2 results show a sharp rise in discriminability beyond a BAC level of 0.03 with failure rates above 50 percent

TABLE B-1. AVERAGE OF n TRIALS TEST STRATEGY APPLIED TO DUNLAP DATA BASE FOR α = 0.025

BAC RANGE	.0 .015	.015 .035	.035 .065	.065 .075	.075 .112	.112 .127	.127 .500	
EVENTS FAILURES % FAIL	1368 51 3.7	84 2 2.4	285 24 8.4	45 6 13.3	348 95 27.3	96 55 57.3	366 273 74.6	Avg. 2
EVENTS FAILURES % FAIL	912 40 4.4	56 1 1.8	190 20 10.5	30 7 23.3	232 80 34.5	64 45 70.3	244 201 82.4	Avg. 3
EVENTS FAILURES % FAIL	456 24 5.3	28 1 3.6	95 16 16.8	15 3 20.0	116 45 38.8	32 22 68.8	122 111 91.0	Avg. 4

n	K_{σ_n}	P_F
1	1.96	.025
2	1.39	.0823
3	1.13	.129
4	.98	.164

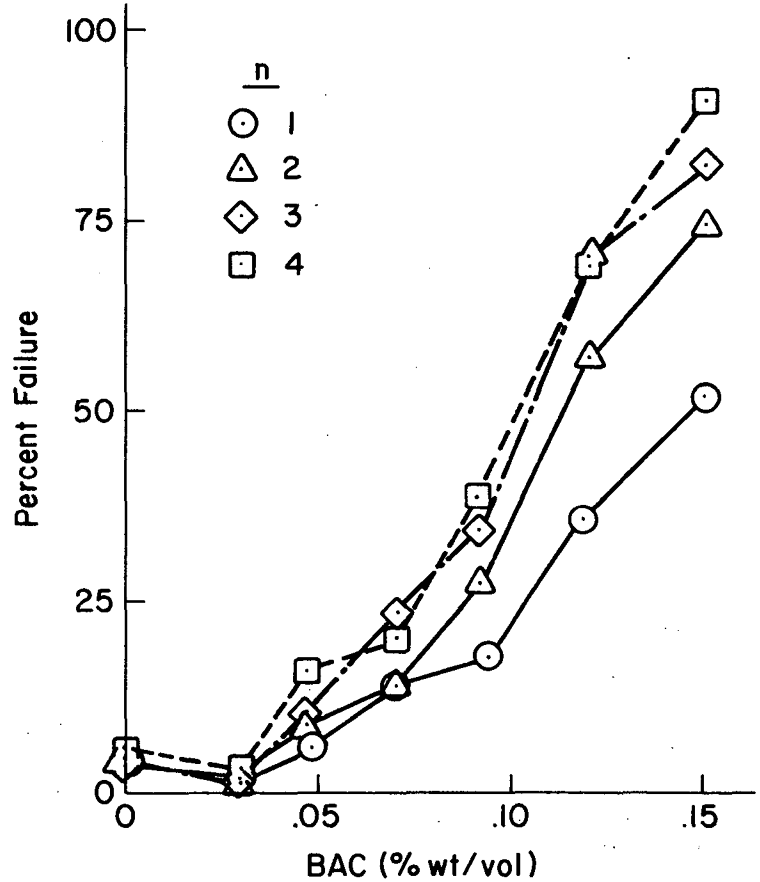


Figure B-2. Discriminability of Average of n Trials Strategy Applied to Dunlap Data Base for α = 0.025

TABLE B-2. 1 OUT OF n TRIALS TEST STRATEGY (1/n) APPLIED TO DUNLAP DATA BASE FOR $\alpha = 0.025$

BAC RANGE	.0 .015	.015 .035	.035 .065	.065 .075	.075 .112	.112 .127	.127 .500	
EVENTS	2736	168	570	90	696	192	732	1 of 1 trial
FAILURES	83	3	38	12	120	69	378	
% FAIL	3.0	1.8	6.7	13.3	17.2	35.9	51.6	
TBARF	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
TBARP	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
TBAR	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
EVENTS	1368	84	285	45	348	96	366	1 of 2 trials
FAILURES	38	1	18	5	82	53	254	
% FAIL	2.8	1.2	6.3	11.1	23.6	55.2	69.4	
TBARF	2.0	2.0	2.0	2.0	2.0	2.0	2.0	
TBARP	1.1	1.2	1.2	1.2	1.3	1.4	1.4	
TBAR	1.2	1.2	1.2	1.3	1.4	1.7	1.8	
EVENTS	912	56	190	30	232	64	244	1 of 3 trials
FAILURES	22	3	19	6	64	41	189	
% FAIL	2.4	5.4	10.0	20.0	27.6	64.1	77.5	
TBARF	3.0	3.0	3.0	3.0	3.0	3.0	3.0	
TBARP	1.3	1.4	1.4	1.6	1.5	1.5	2.0	
TBAR	1.3	1.4	1.5	1.9	1.9	2.5	2.8	
EVENTS	456	28	95	15	116	32	122	1 of 4 trials
FAILURES	14	2	10	2	36	21	99	
% FAIL	3.1	7.1	10.5	13.3	31.0	65.6	81.1	
TBARF	4.0	4.0	4.0	4.0	4.0	4.0	4.0	
TBARP	1.5	1.7	1.6	1.8	2.1	2.9	2.7	
TBAR	1.6	1.9	1.9	2.1	2.7	3.6	3.8	

TBARF = Average number of trials for a fail.
TBARP = Average number of trials for a pass.
TBAR = Average number of trials for a decision.

n	P_F	K_{σ_n}	$\bar{n}_p$ at BAC = 0
1	.025	1.96	1
2	.158	1.00	1.2
3	.292	.55	1.4
4	.398	.26	1.7

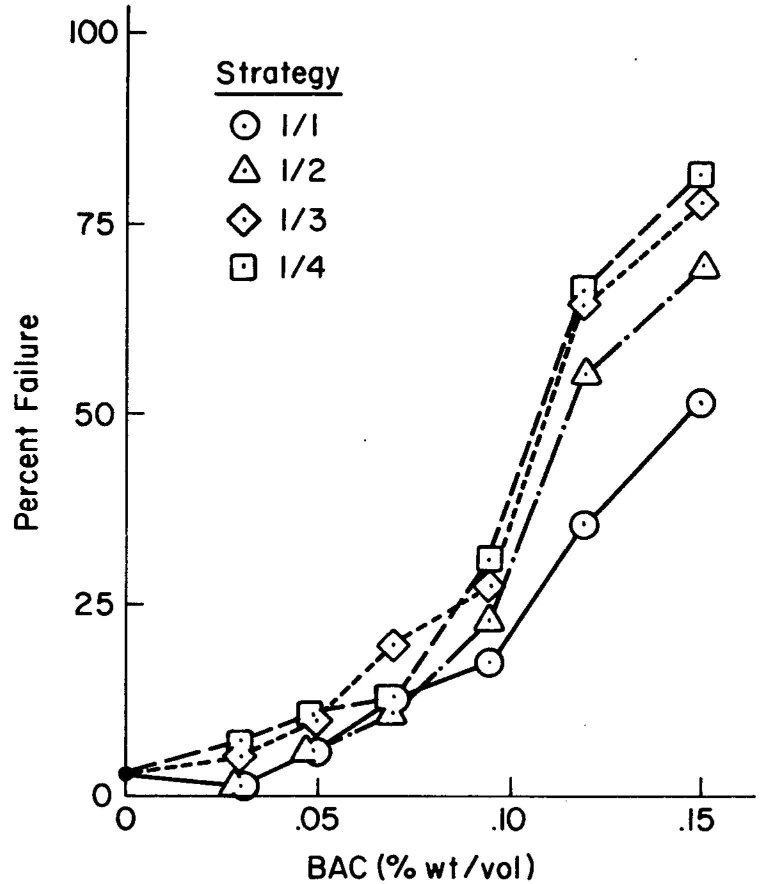


Figure B-3. Discriminability of One Out of n Trials (1/n) Strategy Applied to Dunlap Data Base for α = 0.025

beyond the 0.10 BAC level. There is a distinct improvement in going from 1 to 3 trials, with only minor effects in going to an average of 4 trials.

One Pass Out of n Trials (1/n)

Test failure with this strategy implies failing n trials in a row. For independent trials the probability of n failures is the product of the failure probabilities for each trial

$$P_{F_n} = P_{F_1}^n \tag{B-6}$$

Thus for a specified sober failure rate α, the probability for a single trial failure is set at

$$P_{F_1} = \alpha^{1/n} \tag{B-7}$$

Given P_{F_1} we can then refer to Fig. B-1 or use tables for the cumulative Gaussian distribution function, to determine K_{σ_n}. It is of course possible to pass this strategy in less than n trials, and it can be shown that the average number of trials required for a decision in given by

$$\bar{n} = \frac{1 - \alpha}{1 - P_F} + n\alpha \tag{B-8}$$

The Dunlap data base was again analyzed for the 1 out of n strategy and the results are listed in Table B-2 and plotted in Fig. B-3. Here we note that the desired α level (i.e., percentage of sober failures) was achieved and the predicted mean number of sober trials for a pass was closely matched. The multiple-trial strategies (i.e., $n > 1$) give a definite increase in discrimination of the simple 1 trial strategy (i.e., 1/1). Also, note that the average number of trials for a sober pass is not excessive, even for $n = 4$; however, failing this strategy always requires failing n trials. Discriminability increases appreciably as n is increased from 1 to 3. Diminishing returns are achieved in going to $n = 4$.

Note also in Fig. B-3 that the single trial probability of failure for 4 trials is approximately 40 percent, which is close to each subject's mean and thus a relatively stable region of the CTT score distribution. Referring to Fig. B-1, the pass level λ_p for the 1/4 test strategy can be closely approximated by taking the mean score and subtracting one-tenth of a unit, i.e., for 1/4

$$\lambda_p \cong \bar{\lambda}_c - 0.10 \qquad \text{(B-9)}$$

m Passes Out of n Trials (m/n)

The 1/n strategy is a special case of the m/n strategy. It can be shown that the probability of m passes in n independent trials is given by an expansion series of the negative binomial probability law (Parzen, 1960; Wald, 1962):

$$P_{F_{m/n}} = \sum_{x=1}^{n-m+1} \binom{n}{m-x} P_F^{m-x} (1 - P_F)^{n-m+x} \qquad \text{(B-10)}$$

For several specific strategies, the failure probabilities are given by:

m	n	$P_{F_{m/n}}$
2	3	$P_F^2(3 - 2\,P_F)$
2	4	$P_F^3(4 - 3\,P_F)$
3	4	$1 - (1 - P_F)^3(1 - 3\,P_F)$

(B-11)

To set the pass level a given α is specified, then equations such as given in Eq. B-11 are solved for P_F. Given P_F, Fig. B-1 or Gaussian distribution tables can be used to determine $K_{\sigma_{m/n}}$ as in Eq. B-2. It can also be shown that the average number of trials to reach a decision is given by

$$\bar{n}_d = \left(m + \frac{m\,P_F}{1 - P_F}\right)(1 - \alpha) + n\,\alpha \qquad \text{(B-12)}$$

TABLE B-3. m OUT OF n TRIALS TEST STRATEGY APPLIED TO THE DUNLAP DATA BASE FOR $\alpha = 0.025$

BAC RANGE	.0 .015	.015 .035	.035 .065	.065 .075	.075 .112	.112 .127	.127 .500	
EVENTS	912	56	190	30	232	64	244	2 of 3 trials
FAILURES	30	1	17	8	64	40	180	
% FAIL	3.3	1.8	8.9	26.7	27.6	62.5	73.8	
TBARF	2.9	3.0	2.5	2.6	2.4	2.4	2.2	
TBARP	2.1	2.1	2.2	2.2	2.3	2.5	2.4	
TBAR	2.2	2.2	2.2	2.3	2.3	2.4	2.3	
EVENTS	456	28	95	15	116	32	122	2 of 4 trials
FAILURES	17	1	9	3	38	22	97	
% FAIL	3.7	3.6	9.5	20.0	32.8	68.8	79.5	
TBARF	3.7	4.0	3.6	3.7	3.3	3.3	3.2	
TBARP	2.4	2.5	2.6	2.8	2.7	2.7	3.0	
TBAR	2.4	2.6	2.7	3.0	2.9	3.1	3.1	
EVENTS	456	28	95	15	116	32	122	3 of 4 trials
FAILURES	17	0	9	2	36	21	97	
% FAIL	3.7	0.0	9.5	13.3	31.0	65.6	79.5	
TBARF	3.5	0.0	2.6	2.5	2.7	2.6	2.4	
TBARP	3.2	3.1	3.2	3.4	3.4	3.3	3.6	
TBAR	3.2	3.1	3.1	3.3	3.2	2.8	2.6	

TBARF = Average number of trials for a fail.
TBARP = Average number of trials for a pass.
TBAR = Average number of trials for a decision.

STRATEGY	P_F	$K_{\sigma_{m/n}}$	$\bar{n}_d$ at BAC = 0
1/1	.025	1.96	1
2/3	.09	1.34	2.3
2/4	.245	.69	2.4
3/4	.068	1.49	3.4

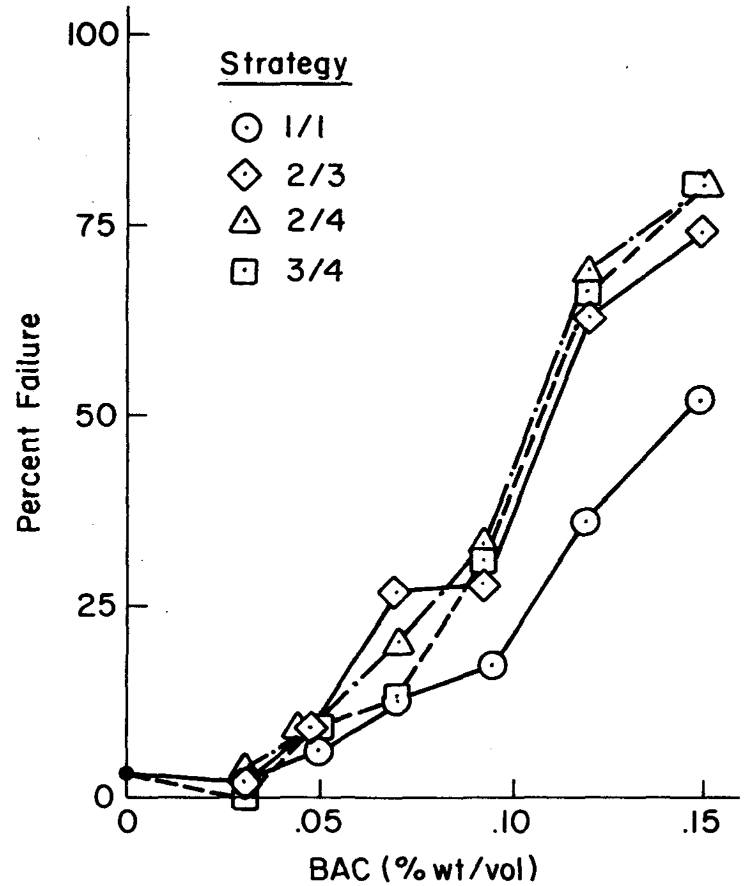


Figure B-4. Discriminability of m out of n Trials Strategy Applied to Dunlap Data Base for $\alpha = 0.025$

The combined Dunlap data base was exercised for the m/n strategy, and the results are listed in Table B-3 and plotted in Fig. B-4. Again we have achieved the desired sober failure rate (approximately 2.5 percent) and have also achieved good discriminability of drunk drivers. The average number of sober trials for a pass has increased over the 1/n case, however, and obviously can never be less than m.

Sequential Strategies With Separate Pass and Fail Levels

Strictly speaking, the 1/n and m/n strategies are sequential tests in that the test is terminated when either the required number of passes or maximum number of fails is reached, and otherwise an other attempt is allowed.

The above strategies only involve one pass/fail performance level, however. We also considered sequential sampling strategies that involve two criteria levels, a high level for passing and a lower level for failure. If a trial results in an intermediate λ_c level, then the driver is allowed to take another test. General sequential strategies are treated in by Wald (1962). Two strategies were considered here as discussed below.

Simple Sequential Strategy. Consider first a simple sequential strategy. Assume on any one trial the probability of passing is P_p, the probability of failing is P_F, and hence the probability of taking another trial is $1 - (P_p + P_F) = 1 - P_T$. Given these definitions we can then make the following table for the test outcome probability after n trials.

TRIAL #	PASS FRACTION	FAIL FRACTION	UNDECIDED FRACTION (take another trial)
1	P_p P_F	$1 - P_T$	
2	$P_p(1 - P_T)$	$P_F(1 - P_T)$	$(1 - P_T)^2$
3	$P_p(1 - P_T)^2$	$P_F(1 - P_T)^2$	$(1 - P_T)^3$
:	:	:	:
n	$P_p(1 - P_T)^{n-1}$	$P_F(1 - P_T)^{n-1}$	$(1 - P_T)^n$

(B-13)

It can be shown that the probability of reaching a decision (i.e., pass or fail) is distributed as a negative binomial distribution (Wald, 1962) and that the mean number of trials to reach a decision is given by

$$\bar{n}_d = P_T^{-1} \tag{B-14}$$

Given that a decision is reached in n trials, then the a priori probability of failing is the proportion of single trial failures relative to the proportion of single trial decisions:

$$P_{F_n} = \frac{P_F}{P_T} \tag{B-15}$$

For a specified Type I error, α, we can solve Eqs. B-14 and B-15 to give the single trial probabilities for passing and failing:

$$\begin{aligned} P_p &= \frac{1 - \alpha}{\bar{n}_d} \\ P_F &= \frac{\alpha}{\bar{n}_d} \end{aligned} \tag{B-16}$$

TABLE B-4. SIMPLE SEQUENTIAL TEST STRATEGY APPLIED TO THE DUNLAP DATA BASE FOR α = 0.025

$\bar{n}_d$		1.2		1.5		1.7		2.0	
BAC	NE	P/F	TRIALS	P/F	TRIALS	P/F	TRIALS	P/F	TRIALS
.0 .0 .015	456	97.8 2.2	1.2 1.0 1.2	97.1 2.9	1.4 1.5 1.4	97.4 2.6	1.7 1.9 1.7	96.7 3.3	1.9 2.9 1.9
.015 .032 .035	28	100.0 0.0	1.2 0.0 1.2	96.4 3.6	1.6 4.0 1.7	96.4 3.6	1.9 4.0 1.9	92.9 7.1	2.0 5.0 2.2
.035 .049 .065	95	96.8 3.2	1.4 2.0 1.4	95.8 4.2	1.7 3.0 1.8	94.7 5.3	1.9 5.6 2.1	94.7 5.3	2.2 5.6 2.3
.065 .069 .075	15	86.7 13.3	1.3 1.0 1.3	86.7 13.3	1.5 1.0 1.5	86.7 13.3	2.2 3.0 2.3	86.7 13.3	2.2 3.0 2.3
.075 .090 .112	116	75.9 24.1	1.4 1.9 1.6	71.6 28.4	2.0 2.2 2.0	66.4 33.6	2.3 2.7 2.5	62.9 37.1	2.7 3.1 2.8
.112 .120 .127	32	50.0 50.0	1.6 1.8 1.7	43.8 56.3	2.7 2.1 2.4	37.5 62.5	3.3 3.3 3.3	34.4 65.6	3.5 3.4 3.4
.127 .134 .599	122	29.5 70.5	1.8 1.6 1.7	21.3 78.7	2.5 1.9 2.1	16.4 83.6	3.3 2.2 2.4	13.9 86.1	3.5 2.4 2.5

NUMBER OF CASES

PASS %
FAIL %

BAC RANGE
High
Mean
Low

MEAN NO. OF TRIALS FOR:
Decision
Fail
Pass

$\overline{n}_d$ at BAC = 0	P_p	K_p	P_F	K_f
1.2	.800	.84	.0205	2.04
1.5	.667	.43	.0171	2.12
1.7	.571	.18	.0146	2.18
2.0	.50	0	.0128	2.23

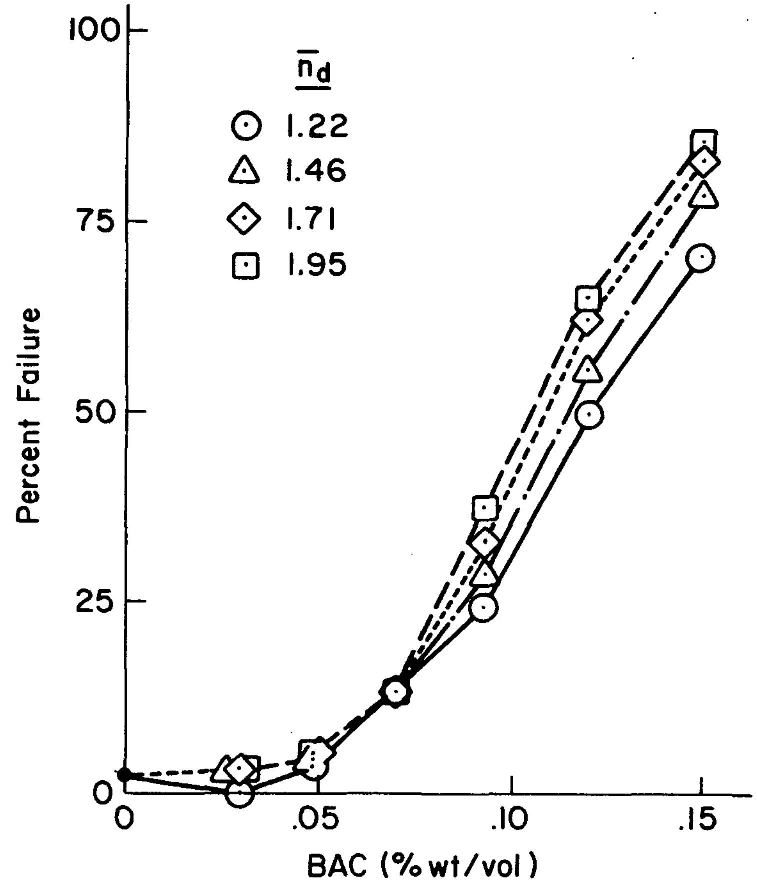


Figure B-5. Discriminability of a Simple Sequential Sampling Strategy Applied to the Dunlap Data Base for $\alpha = 0.025$

Now, by specifying a given α level and selecting desirable values for $\bar{n}_d$, we can then compute pass and fail score levels with the equations

$$\text{Pass Level:} \quad \lambda_P = \bar{\lambda} - K_P\sigma_\lambda$$

$$\text{Fail Level:} \quad \lambda_F = \bar{\lambda} - K_F\sigma_\lambda \tag{B-17}$$

where K_p is chosen to give a one trial probability of passing of P_p and K_F is chosen to give a one trial probability of failing of P_F. Figure B-1 or Gaussian probability tables can be used to determine these levels as discussed previously.

Various values of $\bar{n}_d$ were tested on the combined Dunlap data base. Because each cell of the data base consisted of 6 trials, the maximum number of trials in the simple sequential strategy was truncated to six. If on the sixth trial a decision was not achieved, then a decision was based on the following truncation strategy:

$$\begin{array}{ll} \text{Pass if} & \lambda_c > (\lambda_P + \lambda_F)/2 \\ \text{Fail otherwise} & \end{array} \tag{B-18}$$

The resulting data are given in Table B-4 and plotted in Fig. B-5. Note first that the desired α level (sober failures) was achieved, and the average number of trials for a decision was closely predicted. Note also that the discriminability curves have very desirable characteristics. First, at low BAC levels (i.e., less than 0.075) the failure rate remains low. Also, the various curves are quite consistent at low BAC as compared with the previous strategy plots. This may indicate that the sequential strategy is not as sensitive to anomalies in the data base as the strategies previously discussed. At high BACs (greater than 0.10) the simple sequential strategy shows good discriminability and requires only a modest average number of trials to reach a decision at sober and drunk BAC levels.

Maximum Likelihood Sequential Strategy. A more complicated sequential strategy can be derived from a likelihood ratio test as described in by Wald, 1962. This test involves determining whether the accumulated value of λ_c scores is within a given band:

$$a\sigma^2_{\lambda_1} + (\bar{\lambda}_c + b)n < \sum_{i=1}^{n} \lambda_i < c\sigma^2_{\lambda_1} + (\bar{\lambda}_c + b)n \qquad \text{(B-19)}$$

A score above the band is a pass and below a failure. Another test attempt is permitted for scores within the band. Here we see that the band is offset from the sober performance capability, $\bar{\lambda}_c$, and the size of the band is scaled by the within-subject variance, $\sigma^2_{\lambda_1}$. The coefficients of Eq. B-19 are defined in terms of the Type I (α) and Type II (β) errors and the change in CTT performance ($\Delta\lambda$) it is desired to detect, as follows:

$$a = -\frac{\ln[\alpha/(1-\beta)]}{\Delta\lambda}$$

$$b = \Delta\lambda/2 \qquad \text{(B-20)}$$

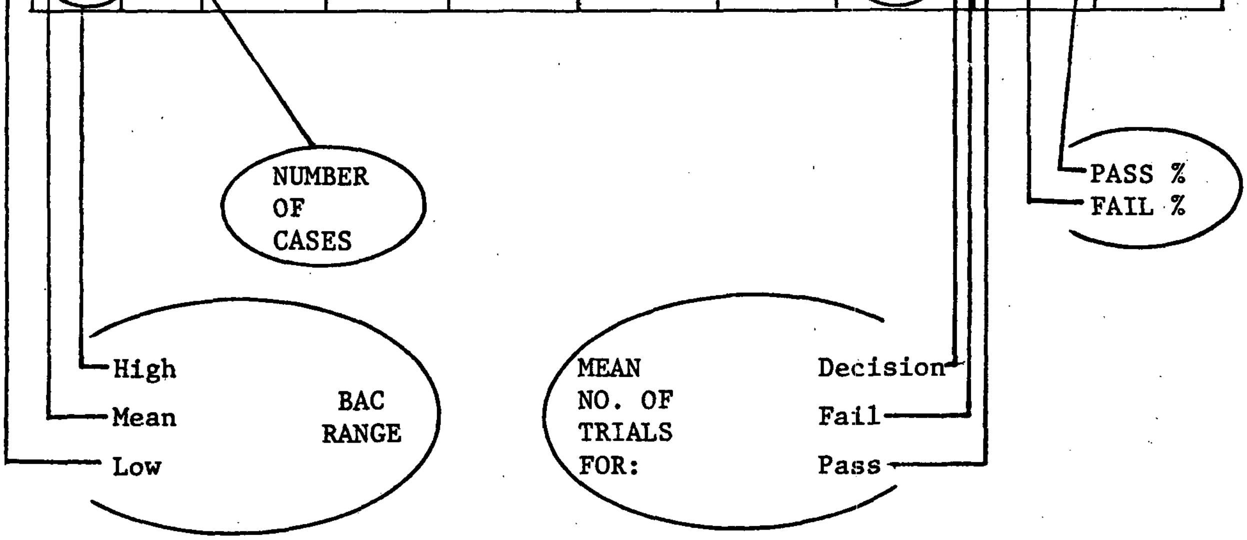
$$c = -\frac{\ln[(1-\alpha)/\beta]}{\Delta\lambda}$$

The combined Dunlap data base was analyzed with the likelihood ratio strategy using individualized means and variances in the Eq. B-20 inequality. Because of the 6 trial limit in the data base cells, a truncation strategy was used where on the sixth trial, if no decision had been reached, the subject was failed if his accumulated score was in the lower half of the band, and passed if it was in the upper half of the band.

The analysis results are tabulated in Table B-5 and plotted in Fig. B-6. The likelihood ratio strategy gives very good discriminability at high BACs (greater than 0.09). At lower BACs (< 0.07) the failure rate is also relatively high for the high discriminability configurations. The likelihood ratio strategy seems to be overly sensitive in the region of 0.05-0.07 BAC as compared with the simple sequential strategy in Fig. B-5.

TABLE B-5. LIKELIHOOD RATIO SEQUENTIAL TEST STRATEGY APPLIED TO THE DUNLAP DATA BASE FOR α = 0.025

CASE		○		△		□		◇	
BAC	NE	P/F	TRIALS	P/F	TRIALS	P/F	TRIALS	P/F	TRIALS
.0 .0 .015	456	98.5 1.5	1.4 2.3 1.4	97.6 2.4	1.4 2.6 1.5	95.2 4.8	1.5 3.6 1.6	94.7 5.3	2.2 4.8 2.3
.015 .031 .035	28	100.0 0.0	1.6 0.0 1.6	96.4 3.6	1.6 4.0 1.6	96.4 3.6	1.7 4.0 1.8	96.4 3.6	2.4 4.0 2.4
.035 .047 .065	95	93.7 6.3	1.8 5.2 2.0	87.4 12.6	1.6 2.5 1.7	85÷3 14.7	1.7 3.3 1.9	85.3 14.7	2.3 5.1 2.8
.065 .070 .075	15	93.3 6.7	1.7 2.0 1.7	86.7 13.3	1.8 4.0 2.1	80.0 20.0	1.6 3.3 1.9	73.3 26.7	3.0 4.0 3.3
.075 .092 .112	116	69.8 30.2	1.9 2.6 2.1	62.1 37.9	1.8 2.6 2.1	53.4 46.6	1.5 2.9 2.1	51.7 48.3	2.8 3.8 3.3
.112 .120 .127	32	40.6 59.4	2.5 2.3 2.3	34.4 65.6	2.4 2.2 2.3	28.1 71.9	2.0 2.3 2.3	18.8 81.3	4.7 3.2 3.5
.127 .149 .500	122	23.0 77.0	2.3 2.1 2.1	18.0 82.0	2.4 2.0 2.1	11.5 88.5	1.9 2.2 2.1	4.9 95.1	3.8 3.0 3.0



CASE	a	b	c	$\bar{n}_d$ at BAC = 0
○	-3.32	-.46	1.12	1.4
△	-3.82	-.35	.62	1.5
□	-4.42	-.23	.02	1.6
◇	-5.76	-.24	1.01	2.3

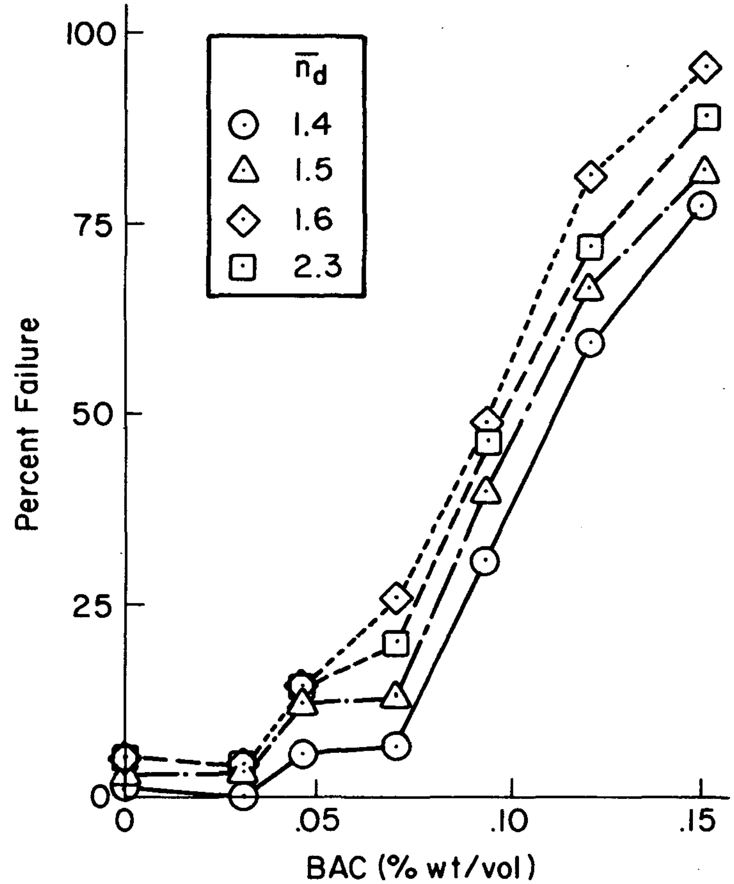


Figure B-6. Discriminability of a Likelihood Ratio Strategy Applied to the Dunlap Data Base for α = 0.025

TEST STRATEGY COMPARISON

The discriminability of several test strategies is compared in Fig. B-7 along with results originally obtained by Dunlap (Oates, 1975b). The results are plotted on probability paper in order to increase the sensitivity to differences at both low and high BACs. The 1 out of 4 and sequential strategies were selected to give a comparable average number of trials for a decision. The likelihood ratio sequential tends to have the best discriminability at high BACs, and the remaining strategies are roughly comparable in the region of 0.10 BAC and above. The strategies studied here are all significantly better than the discriminability results previously obtained by Oates, et al. (1975b). This is not surprising considering they used a universal pass level and a 1 pass out of 2 tries strategy. These differences are analyzed further in Appendix C.

The DDWS device is currently configured to generally handle m pass out of n tries strategies; and short of more exotic decision strategies, such as maximum likelihood procedures, Fig. B-7 shows that the m/n strategy can give adequate discriminability. In particular, the 1/n strategy is desirable in that it allows a low average number of trials for a decision which is of some convenience to the sober driver.

Another desirable property of the 1/n strategy is that for higher n's (i.e., $n > 3$) the individual pass levels approach the subject's mean performance levels. This property leads to two desirable consequences. First, performance is not sensitive to outlying scores (i.e., low probability score variations) in the CTT score distribution. Secondly, an individual's pass level can be estimated more reliably since it is near his/her mean performance level. In comparison, the sequential strategies require fail criterion levels with low single trial failure probabilities. Thus, they would be sensitive to low score outliers in the CTT distribution, and for an individual it would be difficult to reliably estimate fail criterion levels.

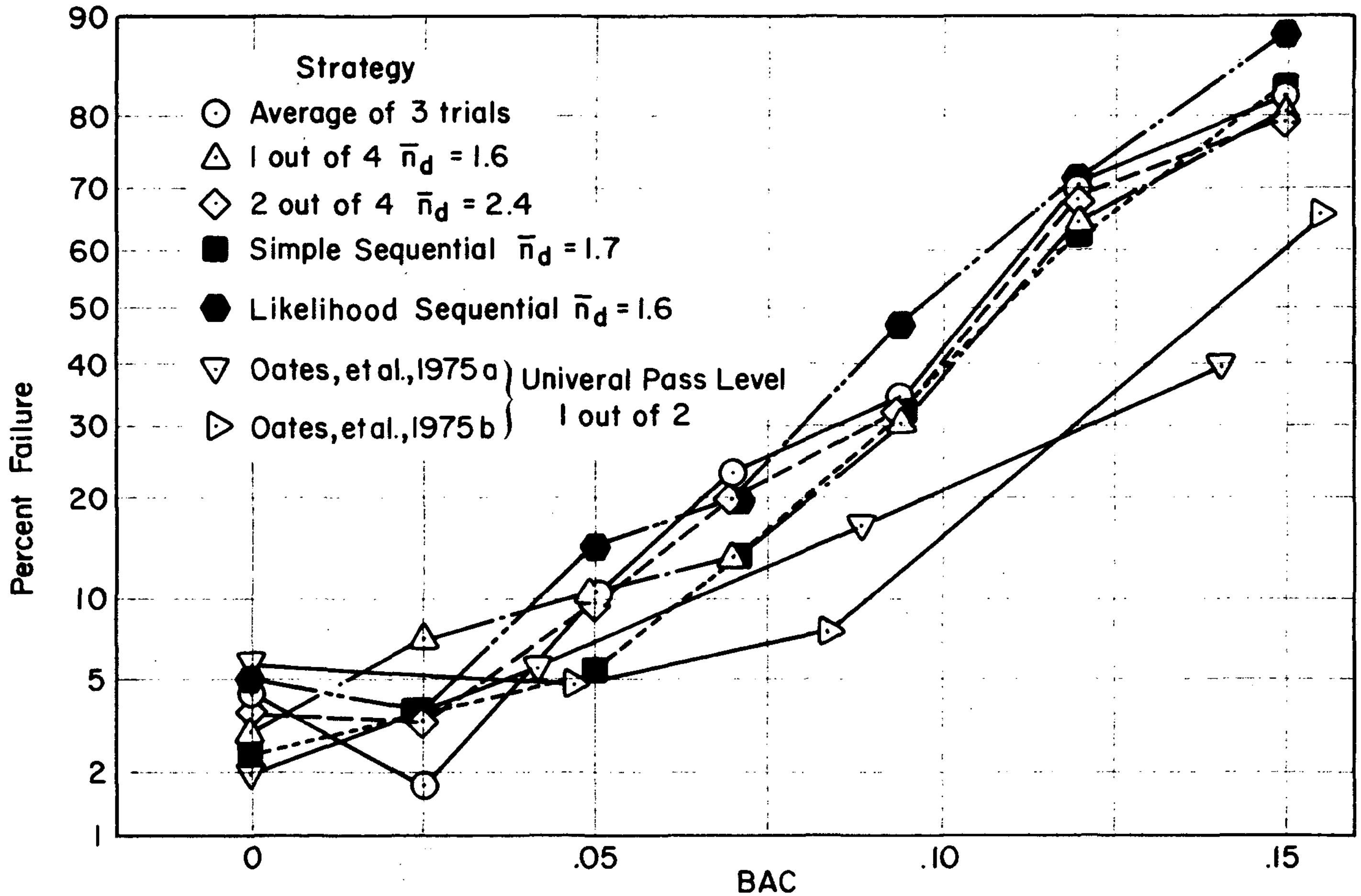


Figure B-7. Comparison of Test Strategy BAC Discriminability

Based on reanalysis of the Dunlap data bases it is thus concluded that the 1/n test strategy has several desirable properties that make it a good compromise selection. For reasonable n's these properties include adequate discriminability, a low average number of sober trials for a decision, and insensitivity to low score outliers. Other aspects of the 1/n strategy are further analyzed in Appendix C.

APPENDIX C

MODEL VALIDATION AND ANALYSIS

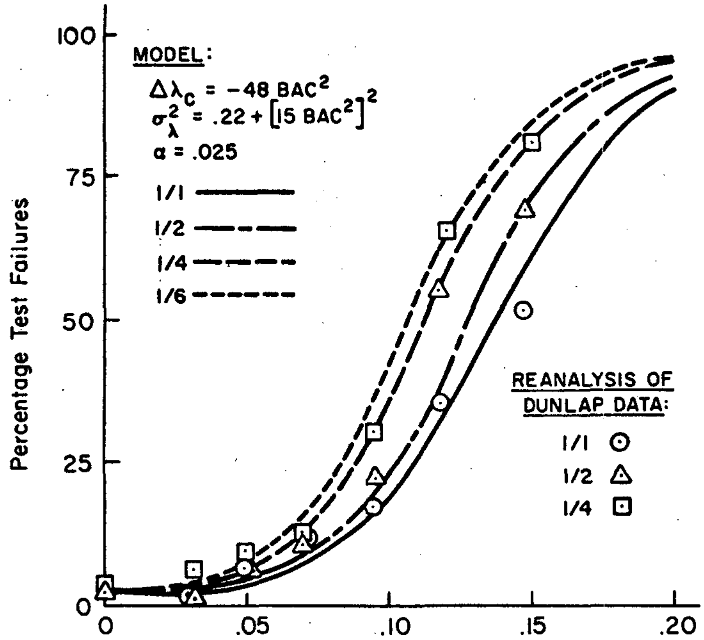
OVERVIEW

In Appendix A a simple statistical model was developed to describe changes in the mean and variability of CTT test scores as a function of BAC. In Appendix B the Dunlap data bases were used to test the BAC discriminability of various test strategies. In this appendix the statistical model is used to analyze variations in test strategy parameters and pass level. Comparisons between model predictions and experimental test data are shown in order to establish the validity of the model.

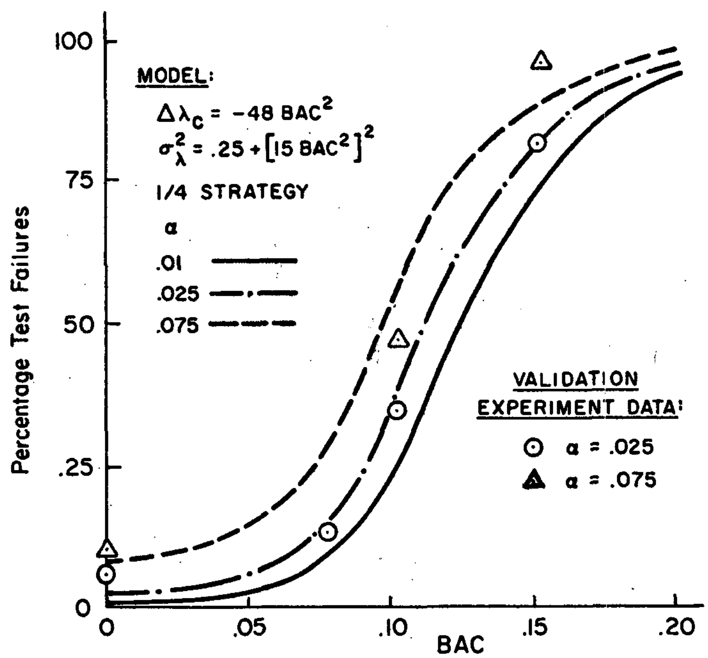
VARIATIONS IN NUMBER OF TRIALS AND SOBER FAILURE RATE

In Appendix B it was concluded that the one out of n (1/n) test strategy offered the best compromise between convenience for sober drivers and discriminability for drunk drivers. In Fig. C-1 the statistical model of Appendix A is analyzed to determine the effects of variations in number of test trials (n) and sober fail rate (α). The model assumes individualized pass levels. The model variance fixed component includes mainly within-subject variability plus some between-subject variability due to unavoidable imprecision in setting the pass levels. The BAC^2 term in the model variance is due to subject-by-alcohol interaction, and the coefficient is based on the BAC regression slope variability analyzed in Appendix A.

In the upper portion of Fig. C-1 we see that for a constant α level (0.025), discriminability increases with the number of trials (n) as expected. Discriminability increases rapidly up to 4 trials, and appears to reach a point of diminishing returns as we increase from 4 to 6 trials. In the lower half of Fig. C-1 discriminability is noted to be quite sensitive to α. Increased discriminability comes at the expense of sober failure rates, however, which is a matter of inconvenience to the sober driver.



a) Variations in n for α = .025



b) Variations in α

Figure C-1. Model Analysis of the Effects of Number of Trials (n) and Sober Failure Rate on the BAC Discriminability of the One of n (1/n) Test Strategy; Model Assumes Individualized Pass Levels

In Fig. C-1 the model curves are noted to be reasonably consistent with experimental data. For the number-of-trials (n) variations in Fig. C-1a the curves have a reasonable match to the Dunlap data base reanalysis of Appendix A. This is not too surprising since the model was derived to match the statistical properties of the Dunlap data. Figure C-1b provides a stronger case for model validity since the comparison experimental data are from the validation experiment described in this report (Section V).

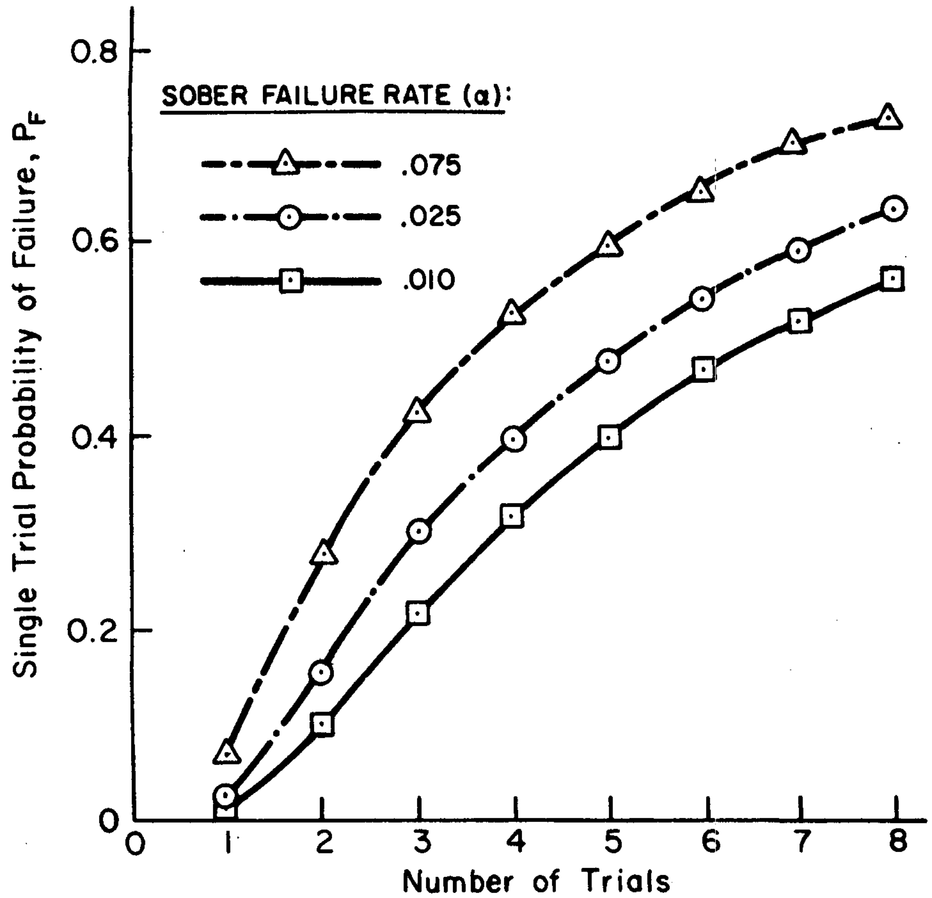
The effect of setting different levels of α on individualized pass levels is illustrated in Fig. C-2. The pass level is basically determined from the one trial failure probability ($\alpha^{1/n}$) shown in Fig. C-2a. The incremental pass level relative to a given subject's mean is then determined assuming $\sigma_\lambda \cong 0.4$ and obtaining a coefficient from a Gaussian cumulative distribution table as discussed in Appendix B. Note in Fig. C-2b that the incremental pass criterion level $\Delta\lambda_p$ begins leveling off in the region of 4 trials. Also, for n = 4, the pass level differences between the three different α levels is only 0.1 units.

The DDWS units can only be set to the nearest 0.1 units, and the estimation of $\Delta\lambda_p$ for a given subject is probably not much better than 0.1 units. Using the 1/n test strategy and determining a given subject's mean or median performance, then subtracting 0.1 units is probably a good means for determining individualized pass levels based on the above analysis. This ensures a nominal sober failure rate ($\alpha \cong 0.025$) and reasonable BAC discriminability.

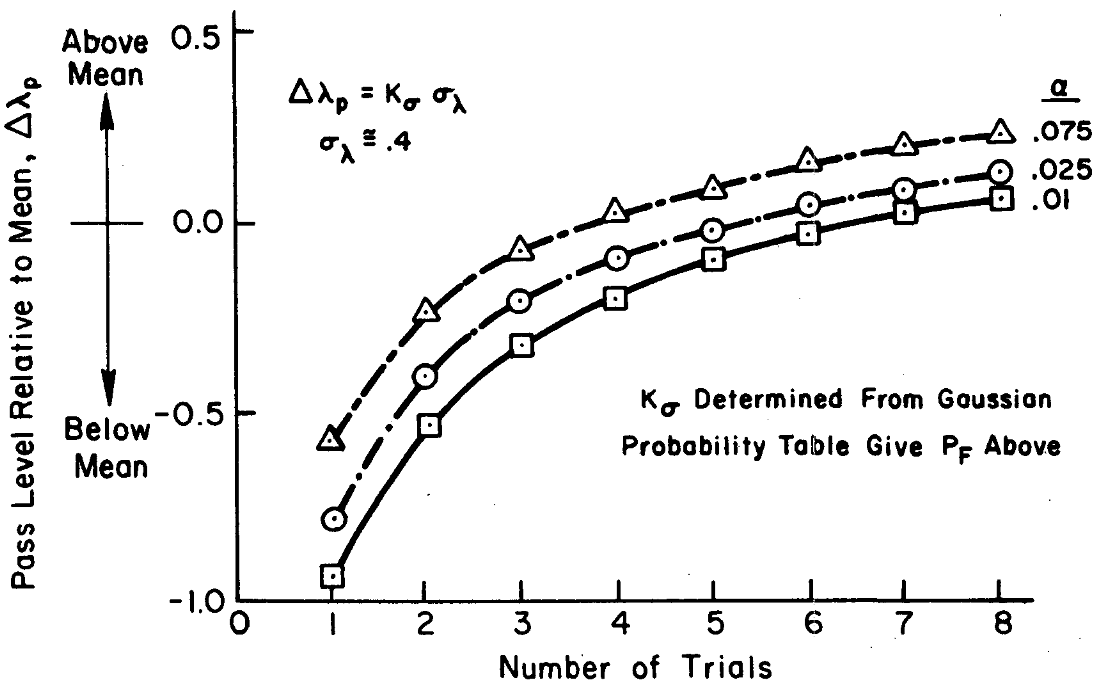
INDIVIDUALIZED VERSUS UNIVERSAL PASS LEVEL CRITERION

In the Dunlap research (Oates, et al., 1975a) it was concluded that a one-out-of-two strategy (1/2) with a universal pass criterion was the optimum test strategy. It is instructive to analyze these Dunlap discriminability results further in order to illustrate the effects of universal vs. individualized pass criteria.

In Table C-1 we have taken the Phase 1 regression analysis model parameters from Table A-2 and computed sober and drunk (BAC = 0.12)



a) Single Trial Probability of Failure ($a^{1/n}$)



b) Incremental Pass Level Relative to Mean

Figure C-2. The Effect of Setting Different Levels of α on Individualized Pass Criteria Levels

TABLE C-1. DISCRIMINABILITY ANALYSIS OF PHASE I DUNLAP DATA (REF. 6) FOR UNIVERSAL VS. INDIVIDUALIZED PASS LEVELS USING REGRESSION MODEL PARAMETERS FROM APPENDIX A

SUBJECT	BAC^2 REGRESSION PARAMETERS			DISCRIMINABILITY CALCULATED FROM REGRESSION MODEL											
				UNIVERSAL CRITERION						INDIVIDUALIZED CRITERION					
				SOBER			BAC = .12			SOBER			BAC = .12		
	a	b	S_r	$K\sigma$	P_F	P_F^2	$K\sigma$	P_F	P_F^2	$K\sigma$	P_F	P_F^2	$K\sigma$	P_F	P_F^2
1	5.22	-106.4	0.697	-1.420	.078	.006	+.778	.782	.612	-.674	.251	.063	1.524	.936	.876
2	4.39	-49.4	0.462	-.346	.365	.133	-.331	.371	.138	-1.017	.154	.024	.522	.700	.490
3	4.43	-107.7	0.460	-.435	.332	.110	2.937	.998	.996	-1.022	.154	.024	2.35	.991	.982
4	4.60	-71.7	0.398	-.930	.176	.031	1.665	.952	.906	-1.181	.119	.014	1.413	.921	.848
5	5.29	-56.2	0.605	-1.752	.040	.002	-.414	.340	.116	-.777	.218	.048	.561	.712	.507
6	4.88	-50.7	0.433	-1.501	.067	.004	.185	.571	.326	-1.09	.138	.019	.60	.726	.527
7	5.18	-35.1	0.480	-1.979	.024	.001	-.926	.178	.032	-.98	.164	.027	.07	.528	.279
8	5.62	-43.8	0.530	-2.62	.020	.000	-1.433	.076	.006	-.89	.187	.035	.30	.618	.382
9	5.17	-48.1	0.512	-1.836	.033	.001	-.483	.316	.100	-.92	.179	.032	.43	.666	.444
10	5.08	-30.9	0.444	-1.914	.028	.001	-.912	.181	.033	-1.06	.145	.021	-.06	.476	.227
11	5.75	-58.3	0.421	-3.610	.000	.000	-1.616	.053	.003	-1.12	.131	.017	.88	.811	.658
12	4.22	-32.0	0.369	.027	.510	.260	1.276	.900	.810	-1.27	.102	.010	-.02	.492	.242
13	4.85	-50.3	0.582	-1.065	.143	.020	.179	.571	.326	-.81	.209	.044	.44	.670	.449
14	4.91	-47.8	0.451	-1.508	.066	.004	.018	.508	.258	-1.04	.149	.022	.48	.684	.468
15	4.55	-36.4	0.406	-.788	.215	.046	.503	.692	.479	-1.16	.123	.015	.13	.552	.305
16	4.84	-37.3	0.480	-1.271	.102	.010	-.152	.440	.194	-.98	.164	.027	.14	.556	.309
17	5.13	-41.3	0.403	-2.233	.013	.000	-.758	.224	.050	-1.17	.121	.015	.31	.622	.387
18	3.79	-29.1	0.396	1.11	.867	.752	2.169	.985	.970	-1.19	.117	.014	-.13	.448	.201
19	4.34	-24.5	0.498	-.221	.413	.171	.488	.688	.473	-.94	.174	.030	-.24	.405	.164
20	4.60	-46.7	0.526	-.703	.233	.054	.575	.712	.507	-.89	.187	.035	.38	.648	.420
21	5.23	-27.6	0.401	-2.494	.006	.000	-1.503	.067	.004	-1.17	.154	.024	-.18	.429	.184
22	4.73	-39.4	0.360	-1.389	.082	.007	.187	.575	.331	-1.31	.152	.023	.27	.606	.367
23	5.83	-56.3	0.485	-3.299	.000	.000	-1.627	.052	.003	-.97	.166	.028	.70	.758	.575
24	4.44	-37.6	0.497	-.423	.337	.114	.667	.750	.563	-1.01	.156	.024	.15	.560	.314
Average Variance	4.88	-48.5 22.9	0.471			.075 .163			.343 .327			.026 .012			.442 .219

failure rates. For a universal pass level in Table C-1 we see that six out of 24 subjects (25 percent) have sober failure rates in excess of 10 percent. Also seven out of twenty-four subjects have drunk failure rates of less than 10 percent. Referring to the discriminability results with individualized pass criteria in Table C-1 we see that the sober and drunk failure rates are much more consistent between subjects. Also, we have been able to select the individualized pass criterion level based on procedures given in Appendix B so that, on the average, we have been able to achieve a lower sober failure rate and higher drunk failure rate.

In Fig. C-3 we have compared the Dunlap discriminability data with the reanalyzed discriminability data discussed in Appendix B. Here it is obvious that the individualized pass criteria allow an increase in discriminability, which is also reflected quite well in the statistical model curves in Fig. C-3.

Another way of interpreting the effect of individualized pass level criteria is illustrated in Fig. C-4. Here, we show typical score distributions as a function of differential CTT scores normalized by the score standard deviation (or normal variate). From the Appendix A Analysis of Variance results note that the standard deviation of within and between subject score variations are approximately equal, and are also roughly equivalent to the average score decrement at 0.10 BAC (i.e., $\sigma_{\varepsilon_1} \cong \sigma_{\varepsilon_2} \cong |\Delta\lambda_{c_{BAC}}| \cong 0.45$). As illustrated in Fig. C-4, 30 percent of the subjects will have average performance levels outside a band of plus or minus one standard deviation unit about the average subject's performance. Setting a typical pass level as a universal pass criterion as illustrated would mean that the poorest performing 15 percent of the subjects would have sober single trial failure probabilities of 85 percent or greater, which translates to a sober failure probability of 50 percent or greater for a one out of four strategy! For the 15 percent best performers, the equivalent single trial failure probability would be 15 percent or less, giving a sober failure probability of 0.05 percent. More pertinent to the 15 percent best performers is that at 0.10 BAC their scores degrade to the average <u>sober</u> subject's score

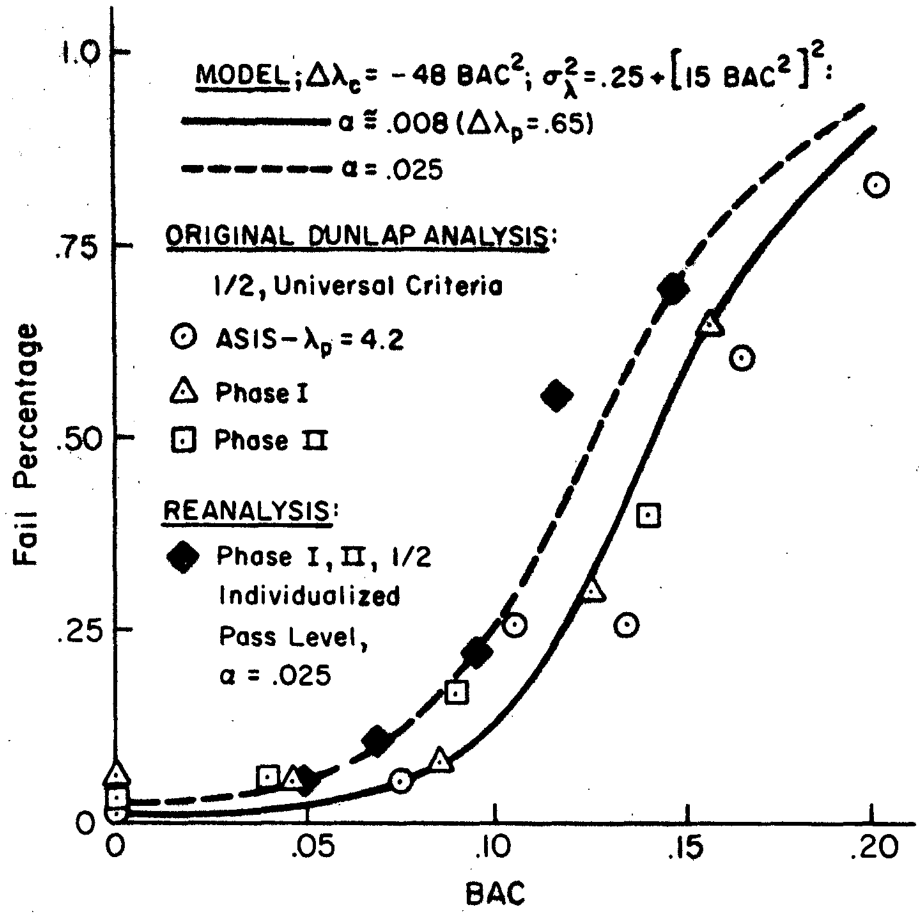


Figure C-3. Discriminability Differences Between Individualized and Universal Pass Criteria Levels

(i.e., approximately a one standard deviation decrement), which still allows them a 50 percent or greater chance of a single trial pass. This translates to a 6 percent chance of failure or 94 percent chance of passing at 0.10 BAC!

The above analysis illustrates two significant reasons for using individualized pass criterion levels. First, the failure rates are more consistent across subjects at a given BAC. Second, a more sensitive discriminability function can be obtained.

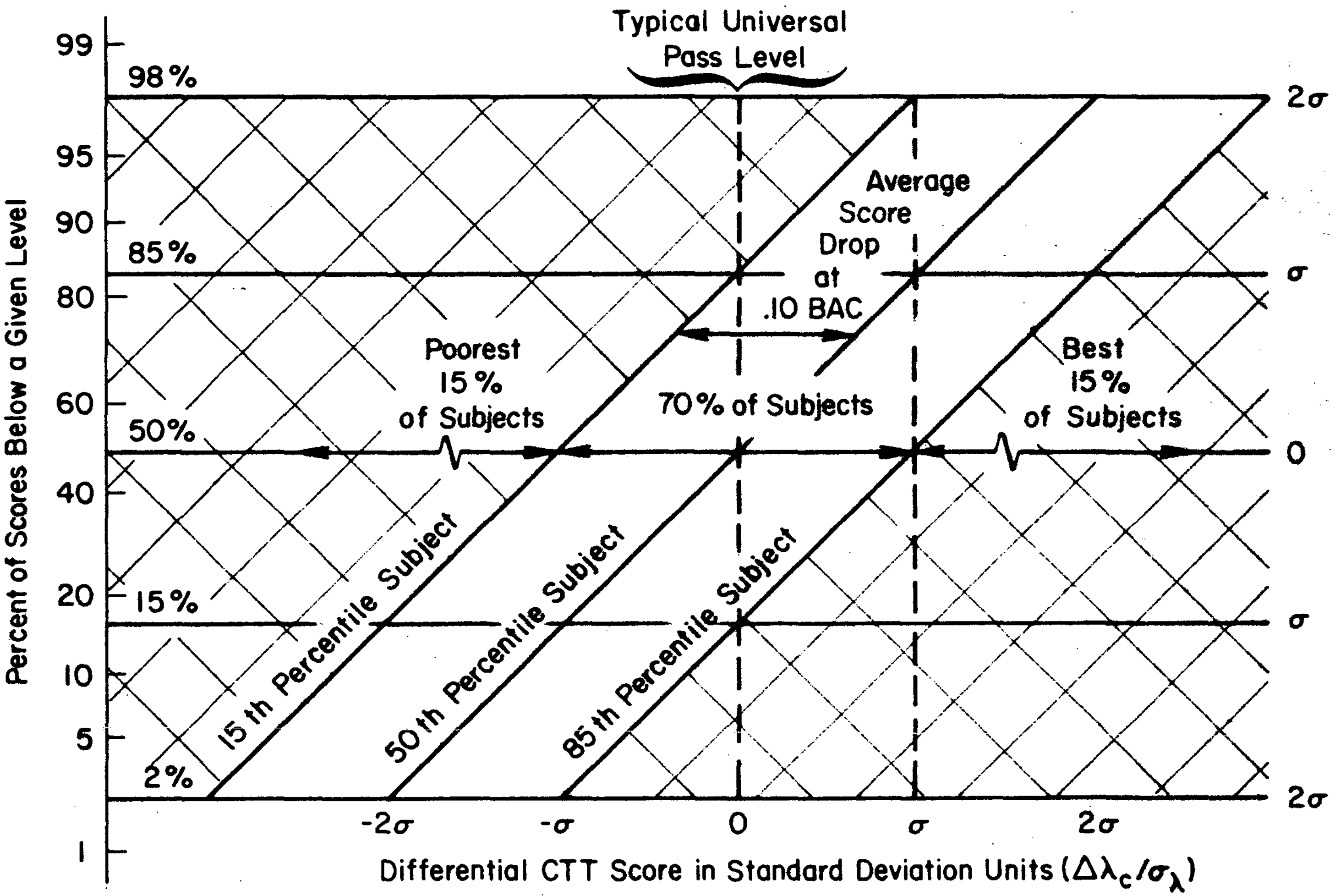


Figure C-4. Hypothetical CTT Score Distributions Based on Appendix A Model Illustrating the Consequence of Universal Pass Level Criteria

APPENDIX D

VALIDATION EXPERIMENT DETAILS

METHODS

Subjects

The subjects included 20 men and 4 women all of whom had been arrested for driving while intoxicated (DWI). Most of the subjects were contacted and enrolled directly through the courts at the time they entered their guilty pleas. A few of the subjects were volunteers who had been contacted through newspaper advertisements. The age range was 22 to 62. All subjects were given the MMPI, and a personal drinking history was taken. The MMPI was used to screen for possible aggression while under the influence of alcohol. The personal drinking history was used to screen out subjects who were not heavy drinkers, in spite of the DWI arrest, and to eliminate anyone with a current history of other drug abuse. Subjects were paid $3.10/hr for all the time they spent on the DDWS project.

FACILITY

Lounge and Drink Mixing Room

The STI alcohol testing facility in Hawthorne consists of a "living room" area, a drink mixing room, and a driving simulator. The living room is arranged to approximate a "real world" drinking atmosphere rather than a laboratory. It is furnished with a couch, chairs, tables, and a TV. It is also supplied with current magazines, playing cards, dominos, a chess set, and crossword puzzle books.

The adjacent room is used for drink mixing and BAC testing. This room contains a refrigerator for storing ice, mixers, and food for lunches. There is a drink mixing table, desk space for the experimenters, and an intoximeter. The intoximeter is a gas chromatograph used to determine blood alcohol concentration from a breath sample. The intoximeter was calibrated daily during the experimental sessions.

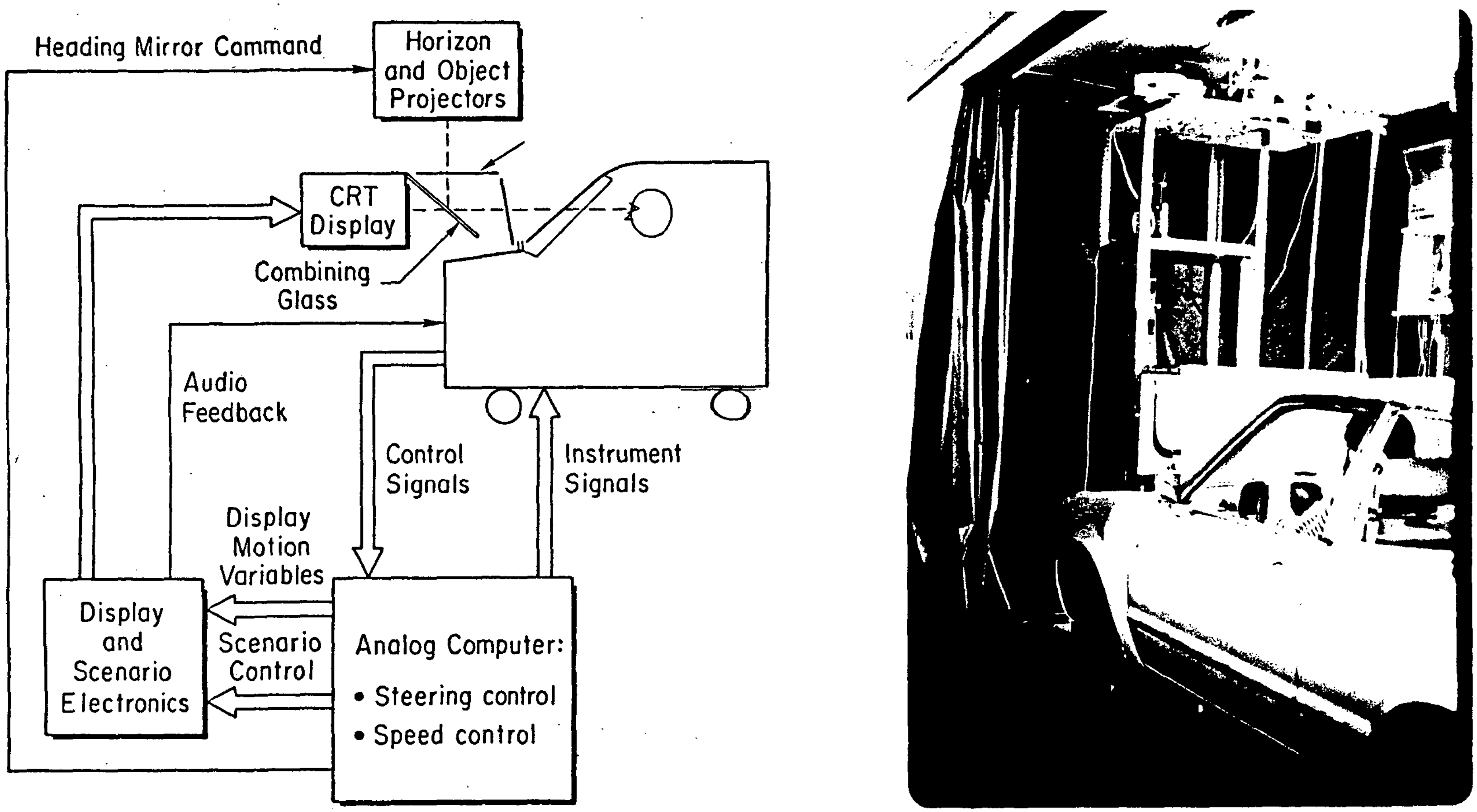
Simulator

STI's interactive driving simulator is composed of an instrumented car cab, analog and digital computers, and a roadway display system. A block diagram of the simulation system is illustrated in Fig. D-1. Control signals from the cab are processed by an analog computer which simulates automobile steering and speed response characteristics. Computed motion quantities were used to drive the cab instruments and roadway display system. The roadway consisted of a cathode ray tube (CRT) presentation of road markings and projected sign images adjusted in size with a computer-controlled zoom lens which simulated approaching signs from a distance. The signs were photographed with 35mm color slide film, then projected in the display generator and optically combined with the delineation background scene. The visual impression was one of a dusk scene, driving under reduced visibility conditions. The background horizon scene was controlled by a servo-driven mirror to provide coordination with the CRT image for car heading changes.

Overall simulator operation was controlled by a modified PDP-11 digital computer. For a more detailed description see Allen and Jex (1980).

DDWS Car (CTT)

The DDWS car was located next to the lounge area just outside of the driving simulator laboratory. The DDWS is installed in a specially modified 1978 Chevrolet Nova. The device consists of a meter (a display and a needle), a start switch, and a small computer in the trunk of the car which mechanizes the Critical Tracking Test and the test strategy. (Complete details are given in Section VI and Appendix F.) When the CTT task begins the needle is centered in the green area of the display; as the task proceeds, the needle begins to wander either to the left or right. The subject tries to keep the needle centered in the green area by moving the steering wheel in the direction he/she wants the needle to move. It becomes increasingly difficult to keep the needle centered and



a) Functional Block Diagram

b) Physical Arrangement

Figure D-1. Interactive Driving Simulator the subject eventually loses control. At the end of the trial, either a green or red light came on signaling to the subject whether or not the pass criterion had been met. The score was copied off a display at the rear of the car by the experimenter and then the CTT was reset for the next trial.

PROCEDURE

There were two components of the subjects' participation in the DDWS research project -- training and the experimental alcohol sessions. The training consisted of three sessions on three different days and included training on both the CTT and the simulator. Usually several subjects were scheduled at one time so they could alternate between the simulator and the CTT. Figure D-2 presents a flow chart illustrating a typical subject's participation, and Fig. D-3 illustrates a typical experimental day.

CTT Training

Each of the 3 training sessions consisted of 25 blocks of 4 trials, for a total of 100 trials. The initial CTT pass criterion was set at an arbitrary low level that all subjects could pass within a few trials (i.e., $\lambda = 2.9$). If the subject passed 3 out of 4 trials in a block, the criterion was raised 0.1. If the subject passed all 4 trials the criterion was raised 0.2. If the subject failed all 4 trials the criterion was lowered 0.1. The pass level could go back to, but not below, a level where 3 out of 4 trials had previously been passed. Breaks were given every 10 or 15 minutes.

The means of the 25 blocks were computed for the last training session. The subject's individual pass criterion for the experimental session was determined by taking the median of the highest 3 block means and reducing the value by 0.3 rad/sec as discussed in Appendix E. This criterion was used for all 3 experimental sessions and did not change.

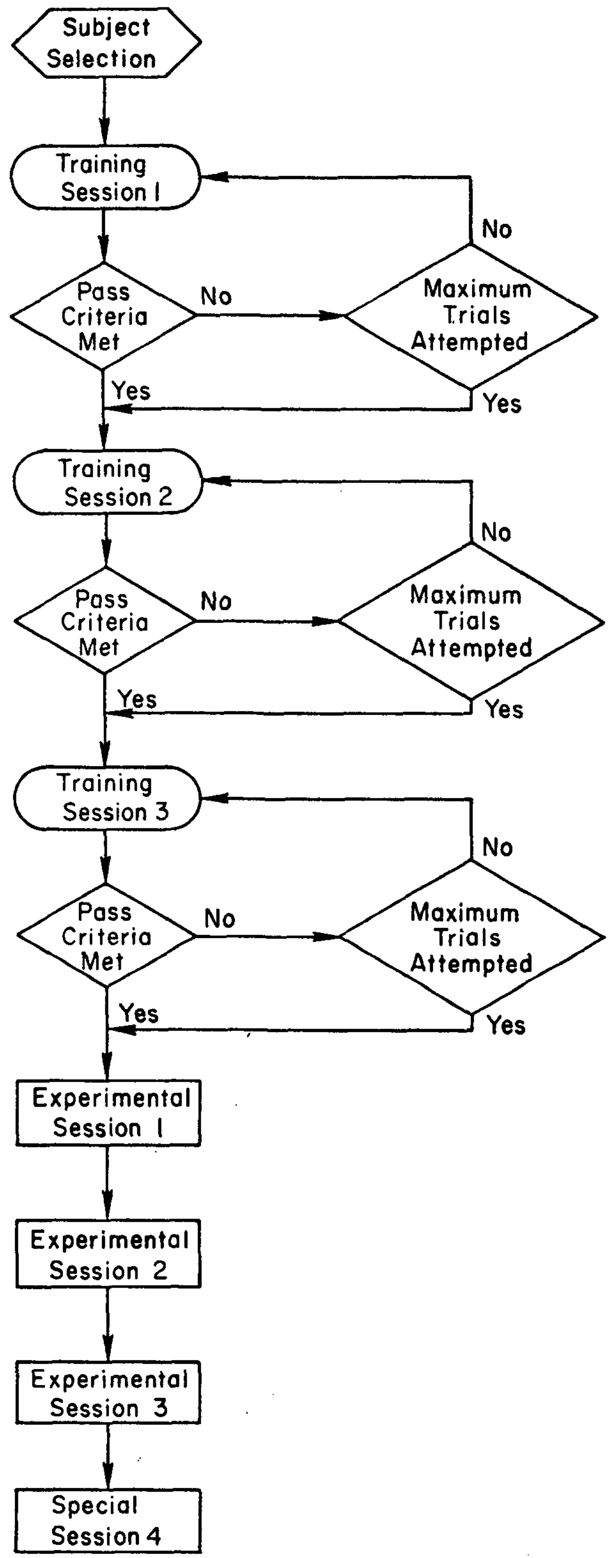
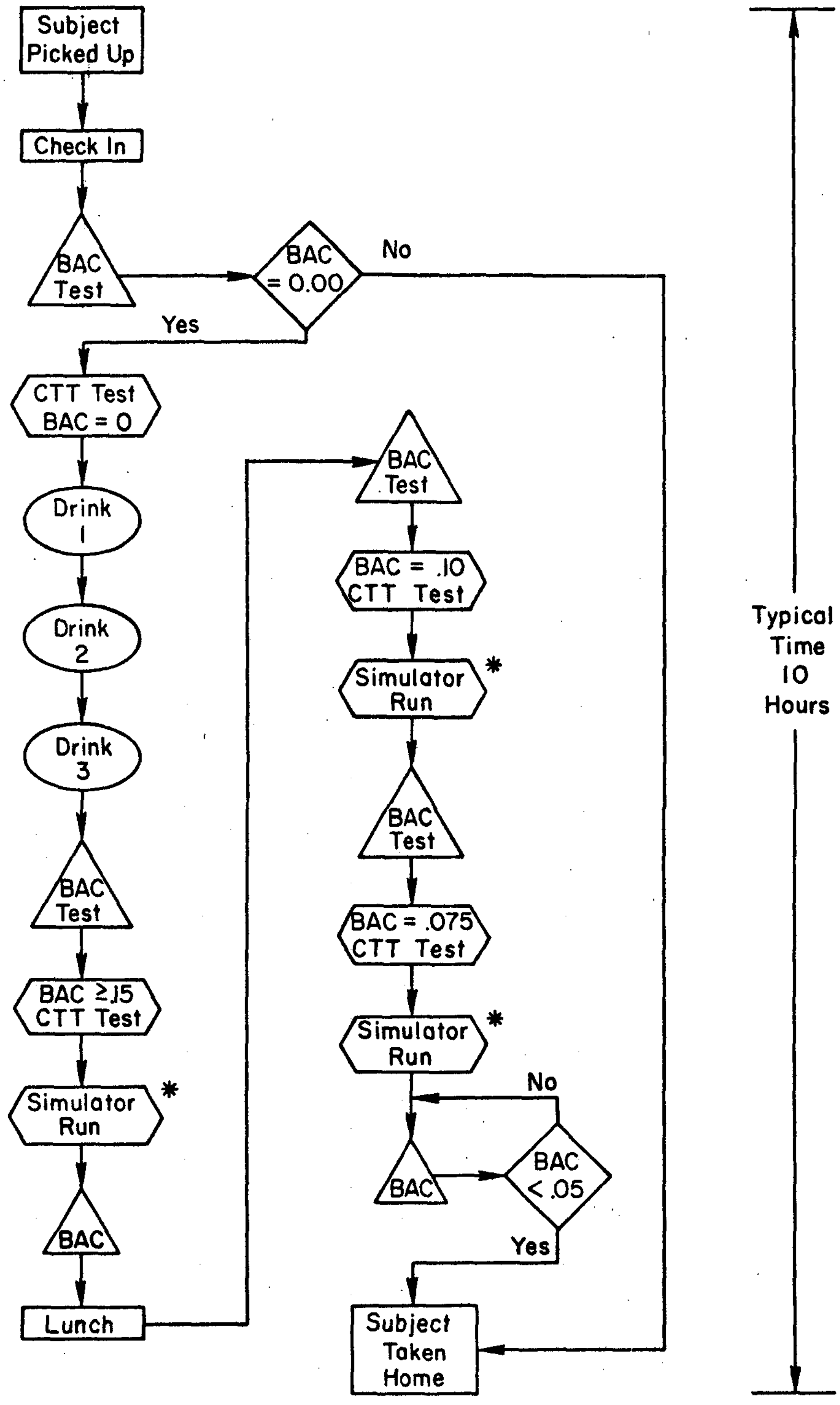


Figure D-2. Typical Subject's Participation



** S does two simulator runs during a session. One of these is elimated according to the experimental design.*

Figure D-3. Typical Experimental Day

Simulator Training

The subjects were oriented as to the objectives of the experiment and the nature of the experimental task, and warned about possible hazards or discomforts. During each of the 3 training sessions, each subject completed 2 simulator runs.

A simulator run was a 15-20 minute drive over a course with a speed limit of 55 mph. During the run the driver was presented with unexpected obstacles, fixed obstacles and curves. Upon encountering an obstacle or a curve, subjects continued around it as safely as possible. By exceeding the roadway boundaries or hitting an obstacle caused an "accident". By exceeding the speed limit, subjects risked getting a ticket from the computer "cop," a siren which was in operation approximately 30 percent of the time.

Subjects also encountered a section of "winding road" with "gusty winds." During this portion of roadway, a series of road signs was presented. The driver was required to respond to the signs by moving the turn indicator, honking the horn, or depressing the dimmer switch.

Subjects were rewarded with a run completion bonus of $10.00. This completion bonus simulated the real world motivation for arriving at a destination. The subject earned $1.00 for each minute saved under 15 minutes. They were charged $1.00 for each minute over 15. They lost $2.00 for each accident and $1.00 for each ticket. Every incorrect roadsign response cost them $.50. Bonus money earned was payable only after completion of the experiment.

REWARD AND PENALTY COMPONENTS

Item	Reward	Penalty
Participation Monies	$ 3.10/hr	
Run Completion Bonus	$10.00	
Time Saved Bonus	$ 1.00/min	
Time Lost Penalty		$1.00/min
Ticket Penalty		$1.00
Sign Response Error		$ .50
Accident		$2.00

Experimental Sessions

Each subject participated in 3 experimental sessions. Two sessions involved drinking, and in a third session the subjects were given placebos. The order of the placebo session was different for each of three subject groups. Otherwise the experimental procedures were the same. Sessions were scheduled about a week apart and the subjects were driven to and from the test site to insure that no one drove under the influence of alcohol. The subjects were instructed not to drink past 10:00 pm the previous night and to have only toast and coffee for breakfast.

The formal sessions began with a breath test using the intoximeter. If the subject tested sober, the baseline CTT test was given. This consisted of 3 blocks of 4 trials with 10 minute breaks between each block. The subject then consumed 3 drinks calculated to bring him/her up to a BAC of 0.15 percent. The drinks consisted of a measured amount of hard liquor (e.g., vodka, bourbon, etc.), based on body weight, and mixer (e.g., orange juice). Mixer was required in order to prepare credible placebos, which consisted of a small amount of liquor floated on top of the subject's preferred mixer. The subject had 40 minutes to consume each drink.

One half hour after finishing the third drink, the subject's BAC was tested. At this point, the subject was usually at 0.15 and the three "peak" test blocks were taken on the CTT. Again, the 3 blocks of 4 trials were separated by 10 minutes. Immediately following the CTT test, the subject did a simulator run. After the simulator run, the subject was given lunch and the BAC level began to drop. As the BAC approached the 0.1 level, the subject did another 3 blocks on the CTT with 10 minute breaks between blocks, and then completed another simulator run. The BAC was continually monitored and another set of 3 blocks on the CTT was given at about 0.075.

During the experimental sessions the subjects earned 75¢ every time they passed at least one trial in a block of 4. This money was kept in a "bank account" until the end of the experiment, at which time it was paid out in a lump sum. Usually there were four, but a minimum of three, subjects for the experimental sessions.

Following the third session the subjects were scheduled to come back for a short follow-up session of 1/2 hour. This consisted of 3 blocks of 4 trials on the CTT. The blocks were separated by 10 minutes. At this point, those subjects who were court-appointed volunteers received their bonus money and a letter of completion for the judge. Those subjects who were volunteers received their bonus money.

Data Reduction

CTT data was combined with simulator data which was then read into an IBM 370 computer. Analysis of Variance procedures were applied to the CTT and simulator data to determine the statistical reliability of measured differences for the experimental design independent variables. Special purpose software was also written to analyze simulator performance data as a function of whether or not subjects had passed or failed the CTT.

CTT discriminability was also analyzed as discussed in Appendices B and C. Because of the training problems discussed in Appendix E, it was decided to analyze the CTT data with pass scores different from those actually used in the experimental sessions in order to obtain discriminability results consistent with each subjects actual performance ability during the experimental sessions. A cumulative distribution plot was obtained for each subjects last training session data. Pass criteria for data reduction purposes were then determined by graphically computing each subjects 40 percent single trial failure probability as discussed in Appendix C. These pass criteria were then used in the validation experiment discriminability analysis.

APPENDIX E

CRITICAL TRACKING TASK TRAINING AND PASS LEVEL SELECTION

OVERVIEW

The Critical Tracking Task (CTT) is a psychomotor test that has proven in the past to allow reliable measurement of human operator performance and has been shown to give sensitive decrements to a variety of stresses. In this project the CTT is being tested as a means of detecting human operator impairment due to alcohol. A pass level is set for each subject, based on that subject's trained asymptotic skill level while sober. It is critical that complete training take place before the individualized pass level is set in order that the impairment can be detected.

There have been three previous Dunlap studies designed to evaluate the CTT as a means of detecting alcohol impairment (Oates, 1973; Oates, et al., 1975a, 1975b). Of these studies the latter two have shown some session order effects in the reanalysis done on this project. This indicates that additional learning took place after the training period ended. Further analysis of CTT training data has shown serious learning problems with some individual subjects. The purpose of this appendix is threefold: 1) develop procedures for maximizing training effectiveness; 2) develop criteria for achieving asymptotic training levels; and 3) develop procedures for selecting individualized CTT score pass levels.

PREVIOUS RESEARCH

There are two critical components to the CTT training procedure. They are: 1) the number of sessions and trials per session; and 2) the incentives and feedback provided to the subjects. The three previous Dunlap studies using the CTT and the DDWS optimization experiment were categorized according to methods used. Table E-1 gives a summary of trials and methods.

TABLE E-1. TRIALS AND REWARDS

STUDY	SUBJECT SELECTION	NO. OF SESSIONS	TOTAL TRIALS	MONEY REWARD	FEEDBACK
Oates, 1973	Volunteer	3	300	None	CTT Score
Oates, et al., 1975a	Volunteer	2	108	25¢/trial over 4.6	CTT Score
Oates, et al., 1975b	Volunteer	2	108	25¢/trial over 4.6	CTT Score
DDWS	Court Assigned	3	300	75¢ for 1 pass out of 4 trials	Pass/Fail

Averaged training data for the three Dunlap experiments are shown in Fig. E-1. The Oates (1973) study shows learning close to asymptote. This study used three training sessions of 100 trials each, which seems to be optimal. The subjects also received CTT Score feedback, which adds a game-like quality to the process. On each trial they know exactly how well they did and they can try to achieve a "personal record" on each trial.

The latter two Dunlap studies in Fig. E-1 showed incomplete training. This was indicated by session order effects which means that learning continued during the experimental session. These studies each used two training sessions with 54 trials per session. The subjects received CTT Score feedback and 25¢ for each trial over 4.6. This seems to indicate that the training was incomplete due to the small number of total training trials.

Based on this research, the DDWS validation study (Section V) used three training sessions of 100 trials each. The feedback information was changed, however. In the DDWS application of the Critical Tracking Task it is important that the subject think that passing the test is an "all or none" process. The subject's feedback consisted of "pass" or "fail," rather than actual scores. This eliminated the game-like quality of the training and introduced motivation problems caused by

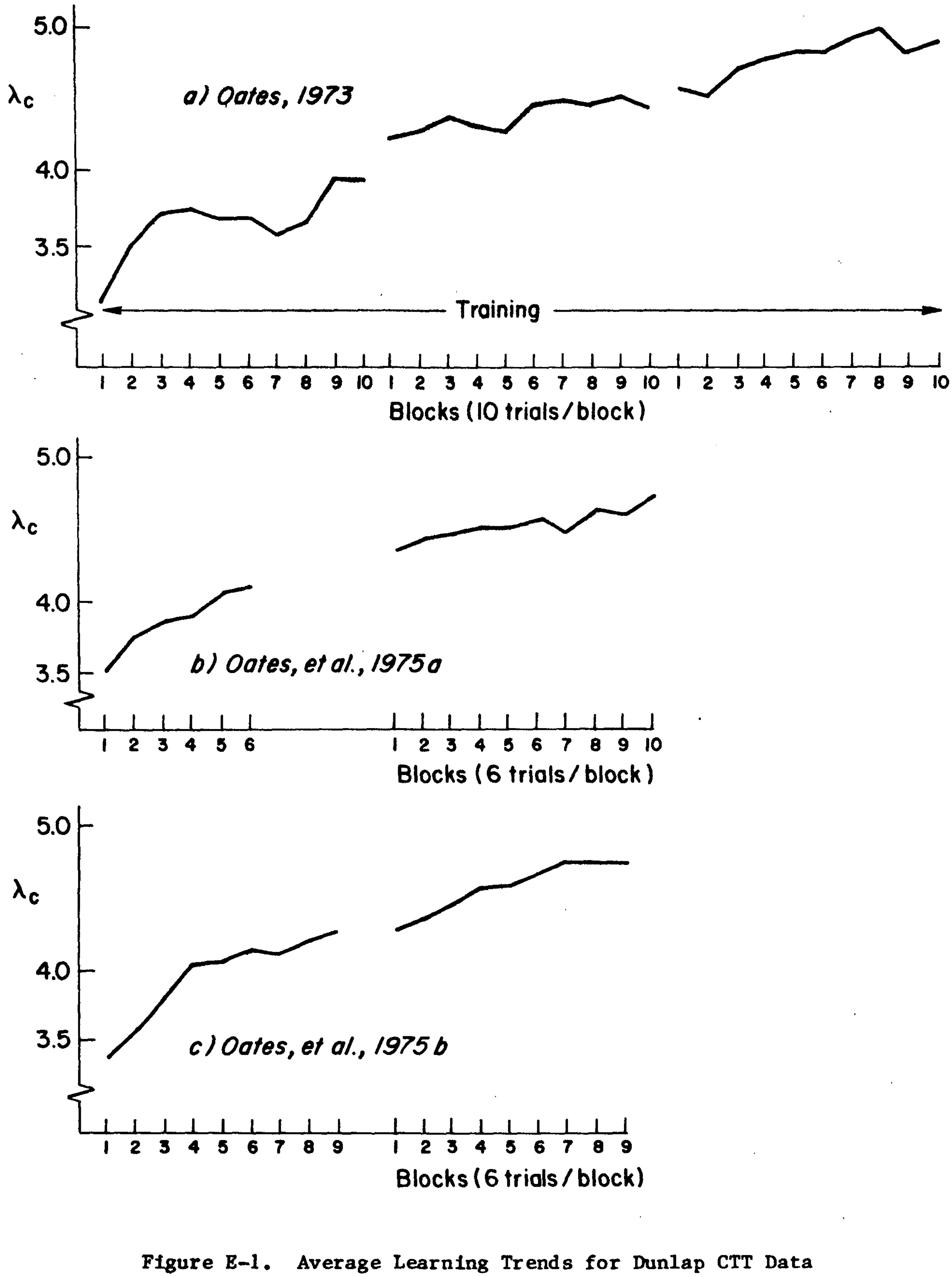


Figure E-1. Average Learning Trends for Dunlap CTT Data boredom. The subjects received 75¢ for passing at least 1 out of 4 trials and could earn up to $30.00 for 1 training session.

Another difference in the DDWS validation experiment was in the basic motivation for participation. In the previous Dunlap studies the subjects were volunteers. In the current study subjects were convicted DWIs who were given a reduced fine for participation in the validation experiment. This means of subject selection was felt to be representative of future DDWS application.

In reanalyzing the Dunlap CTT data the issue of selecting individualized CTT score pass levels was also addressed. It was desired to develop a simple statistical procedure (i.e., one that could be accomplished quickly with a hand calculator) that would minimize the influence of low score outliers typical of training data. For simplicity we worked with 4-trial block means and tried several ad hoc procedures for working with the highest block means on the assumption that these were the most representative of asymptotic training levels.

The best procedure we tried amounted to selecting the second highest block mean as representative of future performance. The accuracy of this procedure is illustrated in Fig. E-2 for the first two Dunlap experiments. Here we see that succeeding experimental session mean CTT scores could be predicted relatively accurately by subtracting a constant from the second highest training block mean. The first Dunlap experiment (Fig. E-2a) shows the least dispersion about the regression line and, referring back to Fig. E-1a, was the experiment with the most complete training.

The Fig. E-2 regression equation for the first Dunlap experiment shows that mean experimental performance can be predicted by subtracting approximately 0.2 units from the second highest training block mean CTT score. Since blocks in that experiment encompassed 10 trials, the second highest block mean can be shown to represent approximately an 84th percentile score with a block standard deviation of $\sigma_\lambda/\sqrt{10}$. Using σ_λ = 0.4 units as the average within-subject standard deviation, the 84th percentile score is approximately 0.13 units above the zero sober

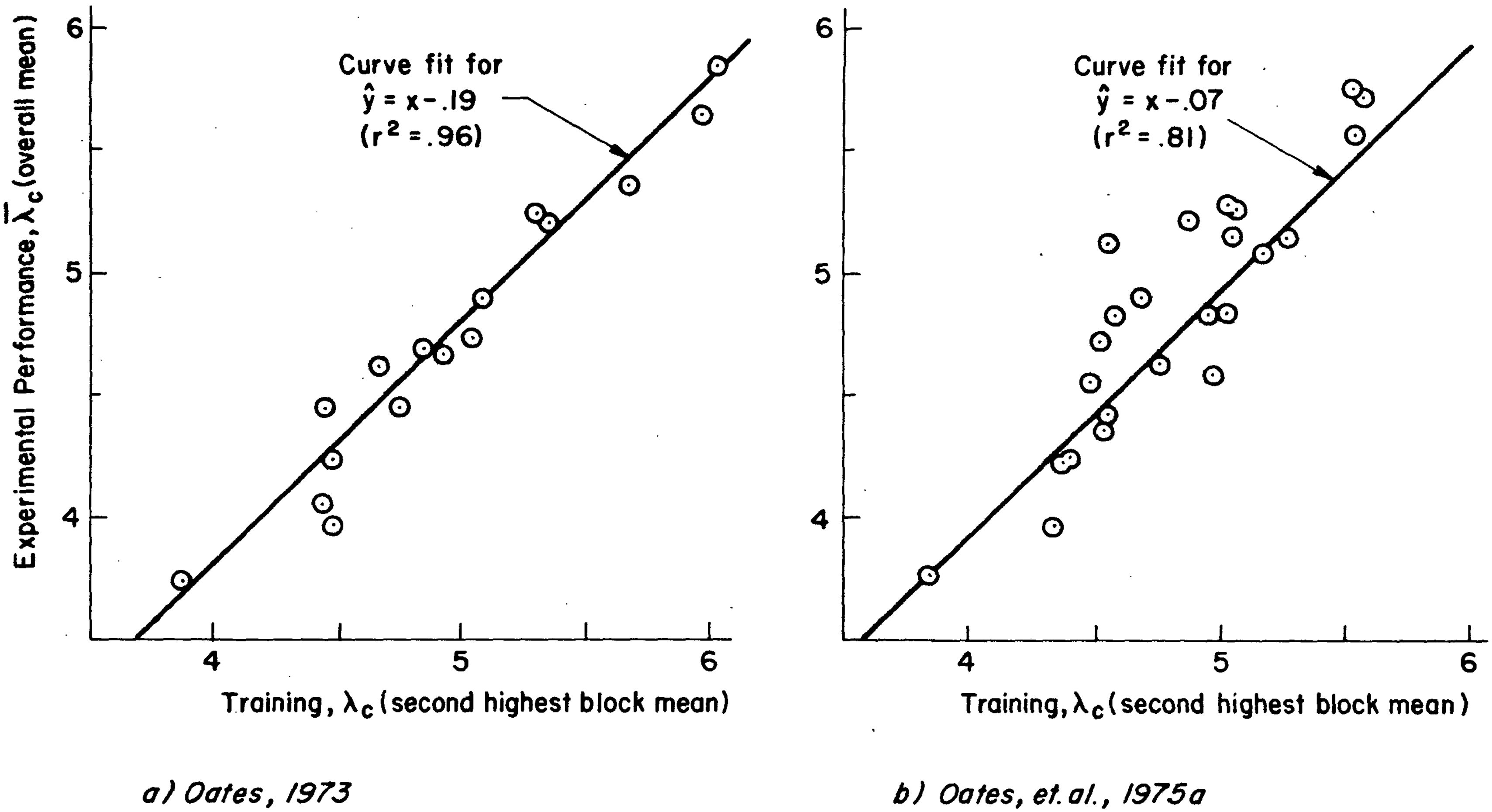


Figure E-2. Comparison of the Second Highest Training Block Mean Score (Last Training Session) with Mean Experimental Performance differential CTT score. Thus, the above simple procedure for determining subject asymptotic CTT performance is consistent with CTT score distributions.

VALIDATION EXPERIMENT

Subjects

For the experiment, 24 subjects who had plead guilty to charges of Driving Under the Influence of Alcohol (DWI or DUI in California) were obtained through the Los Angeles Municipal Courts. The judge offered them the option of participating in this research project instead of the usual $350 traffic fine and traffic school. Even though the monetary rewards for participating were substantial, it was still a choice between the lesser of two evils. Subject participation was often reluctant, and in some cases subjects were effectively participating under duress.

The MMPI was administered and subjects with clinically abnormal profiles were eliminated from the population. The subjects were trained to "drive" in a simulator and to perform the critical tracking task. Three sessions of approximately 2 hours each on separate days were required for training. Following training, subjects participated in three experimental alcohol sessions, scheduled about a week apart, and lasting 10 to 12 hours. Each subject also participated in a follow-up session of approximately 1/2 hour. They received $3.10/hour for training and experimental sessions, and a bonus schedule was worked out to provide incentives for good performance and completion of the program (e.g., Ref. 2). The subjects averaged about $220 total from wages and bonus money and the court cancelled a $350 traffic fine if they satisfactorily completed all requirements.

Learning curves were generated from training data to determine if asymptotic levels of performance had been attained. The overall average learning curve is given in Section V. The 24 training plots were evaluated for indications of poor motivation or inadequate training. Half of the subjects showed lack of motivation and/or not enough training. These curves were characterized by unusually low scores in the last training session and a slope that was increasing even at the end of the last session. The remaining subjects showed asymptotic learning and consistent behavior.

Methods

CTT training was done in 3 sessions on separate days. Each subject read the instructions and was informed of the bonus structure (at least 1 pass in a block of 4 trials = 75¢). Each training session consisted of 25 blocks of 4 trials, or 100 trials.

The initial pass criterion was set at an arbitrary low level that all subjects could pass within a few trials (i.e., $\lambda_p = 2.9$). If the subject passed 3 out of 4 trials in a block, the pass level was raised 0.1. If the subject passed all 4 trials the pass level was raised 0.2. If the subject failed all 4 trials the pass level was lowered 0.1. The pass level could go back to, but not below, a level where 3 out of 4 trials had previously been passed. This strategy tends to prevent deliberate "backsliding" by the subjects.

Breaks were given every 10 or 15 minutes. The mean λ_c for each block in the last training session was computed. The individualized pass levels for the experimental sessions were chosen by taking the second highest block mean and reducing this value by 0.3. This procedure was calculated to give a pass level consistent with a one-trial probability of failure of 40 percent. As discussed in Appendix B, for a one-pass-in-four-attempts decision strategy, this procedure would result in a 2.5 percent sober fail rate.

Results

For this experiment we were interested in determining the extent to which we were able to predict failure rates of impaired operators based on sober training data, and thus needed to predict an accurate pass score (λ_{pass}). The experimental (alcohol) session data (all sober baselines and placebo trials) were analyzed to determine in hindsight what the "ideal" (a posteriori) pass/fail score for each subject should have been to insure a 40 percent probability of failing a single trial when sober. In Fig. E-3 we compare each subject's pass criterion level with the "perfect" level determined in hindsight from analyzing the experimental data. Here we see that subjects identified as well trained from analysis of their learning curves show good agreement. Most subjects with identified training problems were given low pass levels.

Discussion

Because 15 of the 24 subjects were assigned λ_{pass} criteria that were too low, the incentive structure was re-evaluated. The 75¢ bonus had been given for passing at least one trial in a block of 4. (This was chosen because it simulates the one pass out of four attempts strategy used in the formal validation experiment trials in the reanalysis of prior Dunlap data. We found that most of the subjects did not even bother to learn the reward structure in order to maximize the bonuses earned (the bonus money was added across sessions and paid out at the completion of the experiment). They preferred to put forth a random effort and accept whatever total monies they happened to earn.

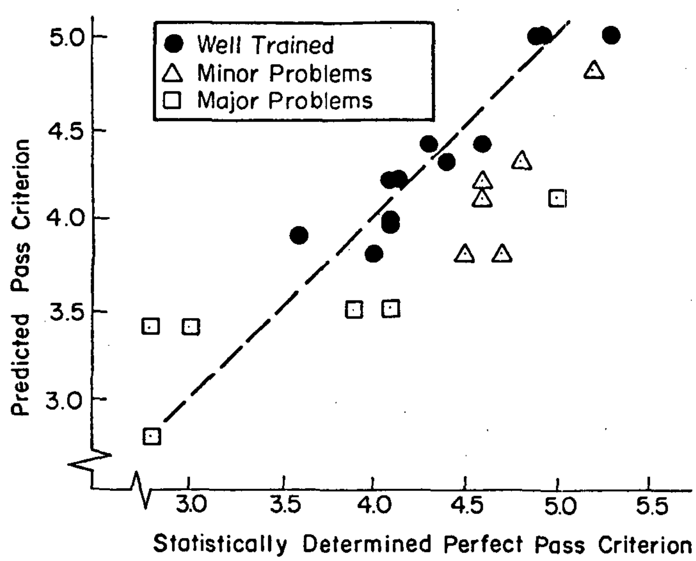


Figure E-3. Pass Criterion Analysis

Research in operant conditioning has shown that the more delayed such a "reinforcer" is, the less potent it is (Reynolds, 1968). Therefore, while the bonus money may have reinforced the subject for completing the experiment, it had little effect on each individual trial of the CTT. It was also observed during training that the subjects' primary motivation was to complete training sessions as soon as possible, even to the extent of foregoing rest breaks and failing the test quickly to speed up the trial repetition rate.

It was decided that in order to elicit a consistent and stable performance from the subject a more immediate reinforcer should be applied after each passed trial on the CTT. Reinforcement occurs when a rewarding stimulus follows a response or when a negative or punishing stimulus is avoided. It was decided that, since the training procedure is so tedious, a negative, or punishing, stimulus would be a 30 second "time out" condition added after each failed trial. By using the avoidance paradigm, the absence of the aversive stimulus (the time out) becomes a reinforcer (Reynolds, 1968).

A subsequent study was conducted to verify if addition of this "time out" procedure would motivate the subjects in order that stable and complete training is obtained in three sessions.

TRAINING EXPERIMENT

Objectives

As described earlier, each subject is required to complete 300 trials on the CTT to establish an individualized pass/fail score. This is a tedious process done over 3 sessions, each lasting 2 to 3 hours.

Because of the poor performance and comments made during the experiment, we conducted a brief training experiment using the reinforcement strategy described above. In this situation a green "PASS" light in the display of the CTT apparatus takes on new reinforcing properties, as it now signals the absence of a "time out." In the previous experiment the green light, in general, only provided information as to the outcome of the trial. The red "FAIL" light, on the other hand, now takes on aversive properties because it is present during the 30 second "time out." The red light was also merely informational in the previous experiment.

Procedure

Six additional subjects were contacted and enrolled through the Los Angeles Courts, as before. One hundred dollars of their $350 fine was dropped when they completed the project. The MMPI was administered as in the earlier experiment. The first two subjects were given the printed instructions that outlined the objectives of the study and the bonus structure. They repeatedly expressed their surprise at being paid, because traffic schools do not pay for participation. Also, as in the earlier experiment, they chose not to learn the incentive structure and said, in effect, "You just keep track and pay me later." It was decided at that point that paying the subjects was superfluous and unnecessary, so the next four subjects received verbal instructions with no mention of wages or incentives.

The subjects received the same training procedure as in the earlier experiment (25 blocks of 4 trials each), except that each time a trial was failed there was a 30 second delay. [Note: These subjects all came after work and, because of the hour and the time of year, the testing, which took place in a parked car, was done in the dark. When the trial was completed, the display light went off with only the red "FAIL" or green "PASS" light remaining on, depending on the outcome of the trial. This meant that for a failed trial the subject waited in a dark car for 30 seconds, looking at a small red light that said "FAIL."]

Results

Some of the learning curves are shown in Figs. E-4 through E-6 for the new condition using the "time out" procedure. There is no qualitative difference between those subjects who were paid and those who were not paid. The learning curves show the same asymptotic learning curves that yielded accurate criteria predictions in the earlier experiment. Furthermore, the trial-to-trial consistency is better, as shown by the low standard deviations at the bottom of Figs. E-4 through E-5.

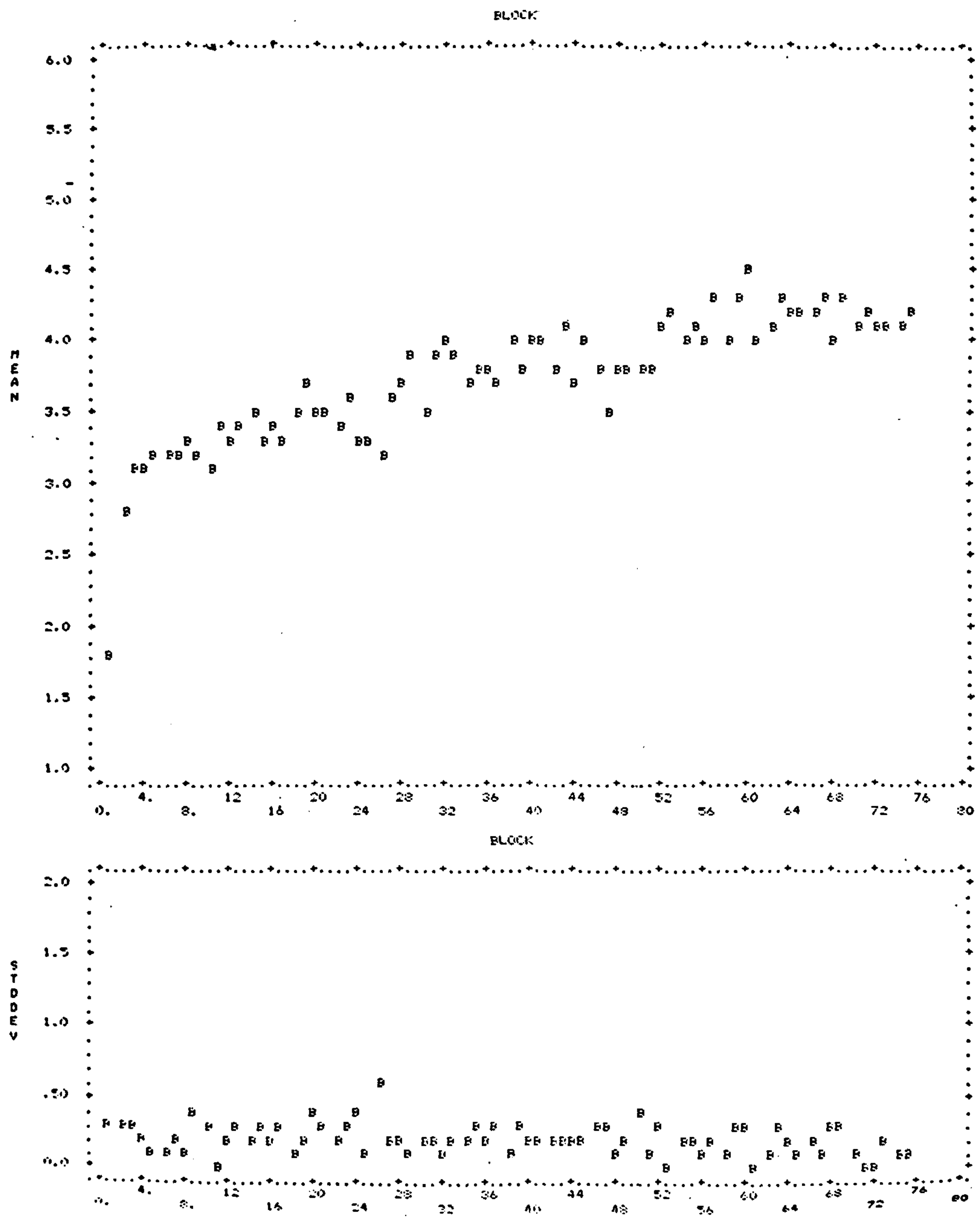


Figure E-4. Learning Curve for a Subject Who Received Delayed Monetary Rewards for Task Performance in Addition to Immediate Wait Time Penalties

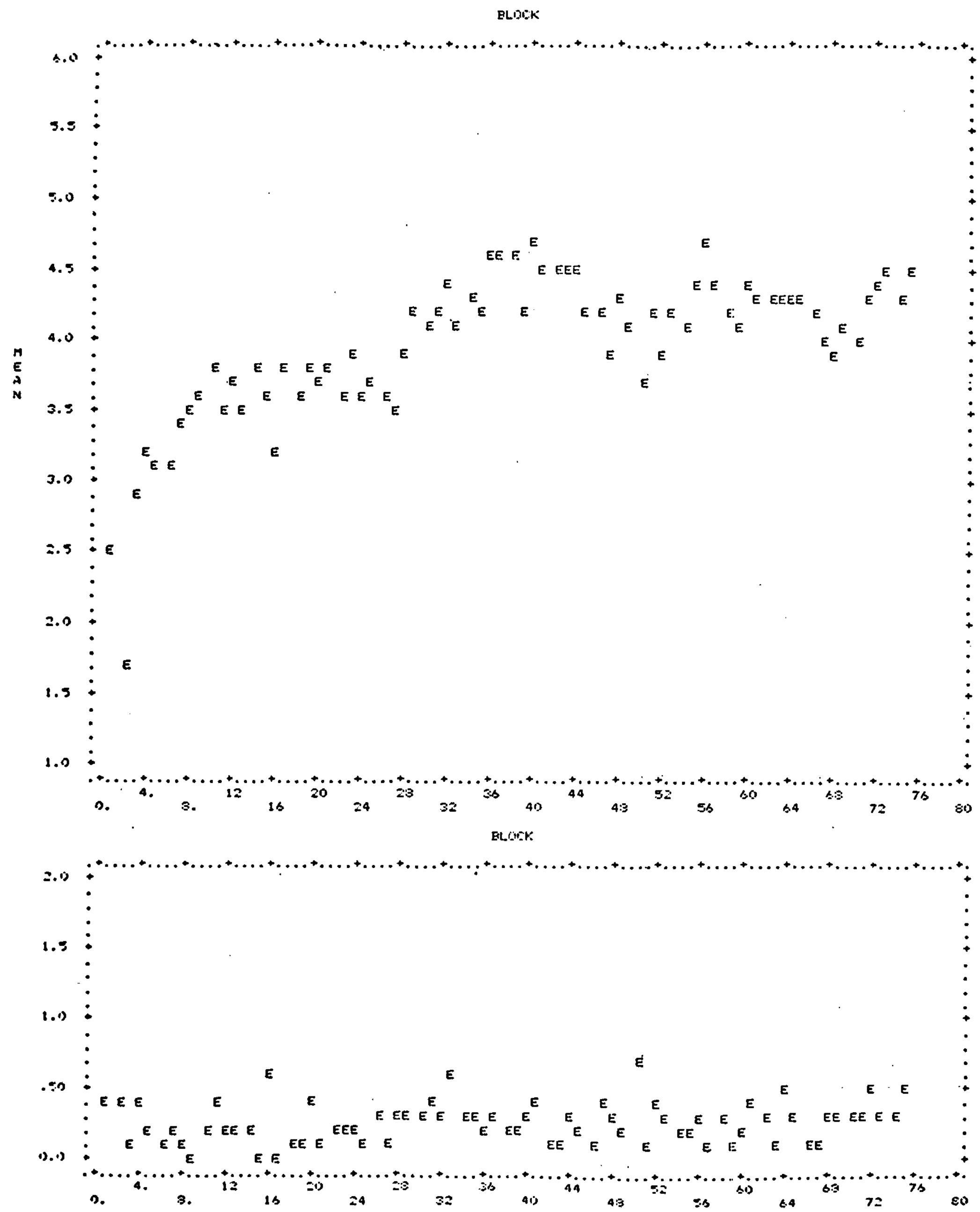


Figure E-5. Learning Curve for a Subject Receiving No Monetary Rewards for Task Performance

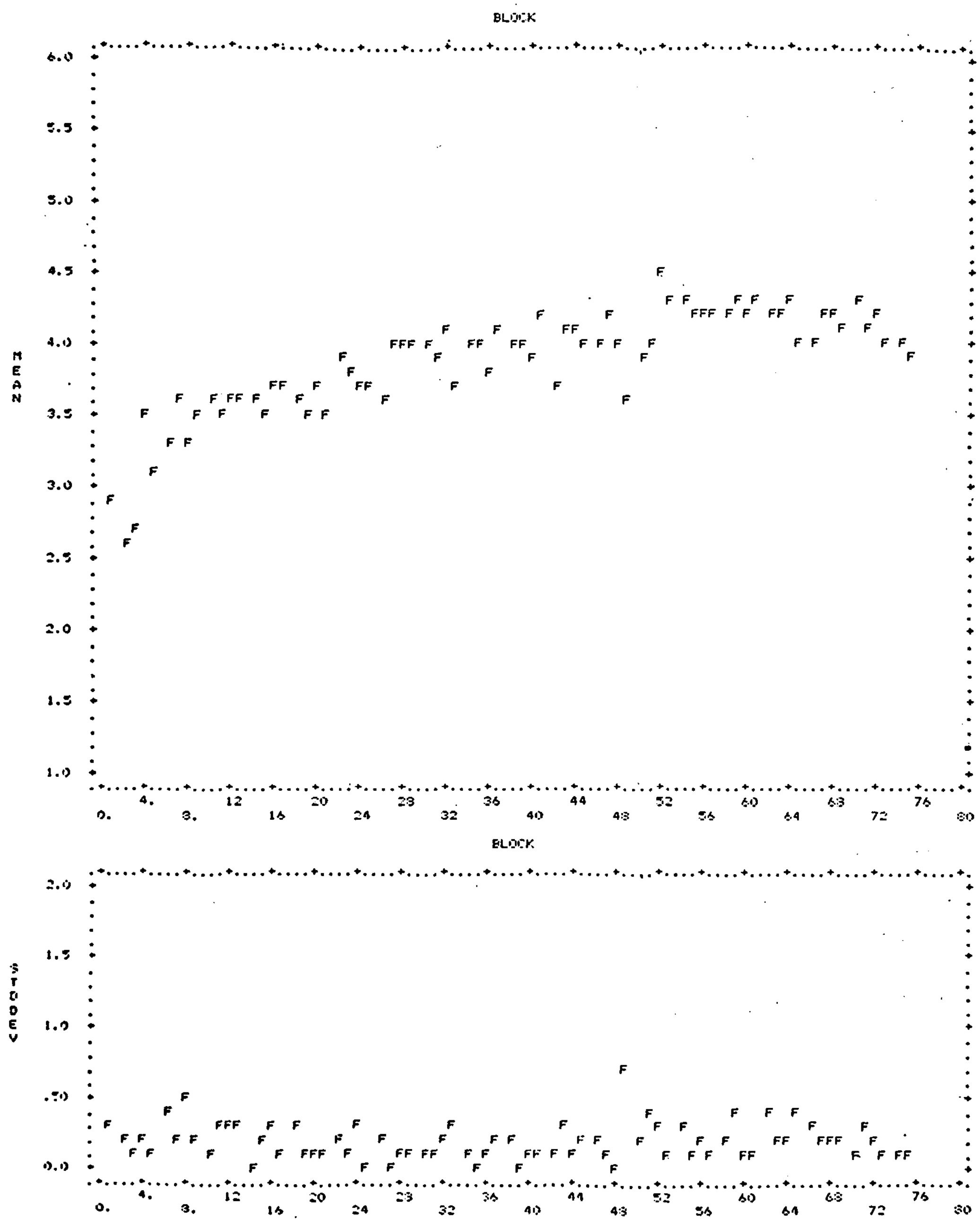


Figure E-6. Learning Curve for a Second Subject Receiving No Monetary Rewards for Task Performance

Discussion

This training study has shown that, to make an accurate prediction of a pass criterion, stable and complete learning must take place. It has also been shown that learning is facilitated when the subject is reinforced after each trial. In the past, subjects were given the CTT Score as feedback on after each trial. This acted as an immediate and powerful reinforcer, because each trial provided an opportunity to better their last score. In addition, subjects in past experiments were motivated volunteers. For these reasons earlier monetary reinforcement structures were based on a positive reinforcement model only. In this experiment the subjects were trained to "pass the test" without being aware that the eventual pass level was adjusted to their individual ability to perform. Since the normal reinforcement by display of CTT scores was not used, the 30 second time out after each fail was substituted.

In assessing what reinforcers are available for use with unmotivated subjects, money seems to be a neutral stimulus. The main motivator should be based on passing the test to avoid the aversive stimulus. In this way subjects will learn to pass the test without being aware that the eventual pass level is adjusted according to their individual ability to perform. The "30 second time out for fail" procedure seems to provide the desired immediate reinforcement after each trial and maintains consistent test scores.

CONCLUSIONS AND RECOMMENDATIONS

- CTT training should be conducted over three sessions with 100 trials per session.
- Start training at a low pass level ($\lambda_p < 3.0$). Increase λ_p 0.1 unit if the subject passes 3 trials in a 4 trial block, 0.2 units if he/she passes all four trials. If all 4 trials are failed reduce λ_p 0.1 units, but do not go below the last 3 passes out of 4 trials level.
- Subjects should be given only pass or fail feedback and a 30 second wait time for each test failure.

- Set the trained CTT Score pass level by taking the second highest 4 trial block mean during the last training session and subtracting 0.3 units.

- As suggested in Appendix C, a second method for setting the pass level criterion is to obtain a subject's median score from a cumulative distribution plot then set the pass level at 0.1 unit below the median.

APPENDIX F

DDWS SYSTEM CONFIGURATION

OVERVIEW

The DDWS system and vehicle were described in general terms in Section VI. This appendix gives more detail in the following areas:

- Physical design
- Functional operation
- Front panel controls
- Data logger

More detailed description of the DDWS electronics can be found in Jex and Peters (1974).

ELECTRONICS

Microcomputer

Figure F-1 derived from Jex and Peters (1974) portrays the system block diagram. The design of the system is based on a microcomputer, the PLS-401 manufactured by the Pro-Log Corp. of Monterey, California. The heart of the microcomputer is a 4-bit microprocessor chip, the 4004 manufactured by Intel. The microcomputer performs the following functions:

- Controls sequencing and timing of tests
- Implements the control task, termed the Critical Tracking Task
- Reads the driver's control input (steering wheel position) during the tracking task
- Provides the signal to drive the task display meter
- Informs the driver of the test status by illuminating one of the three test status lights in the display
- Controls the illumination of the display meter

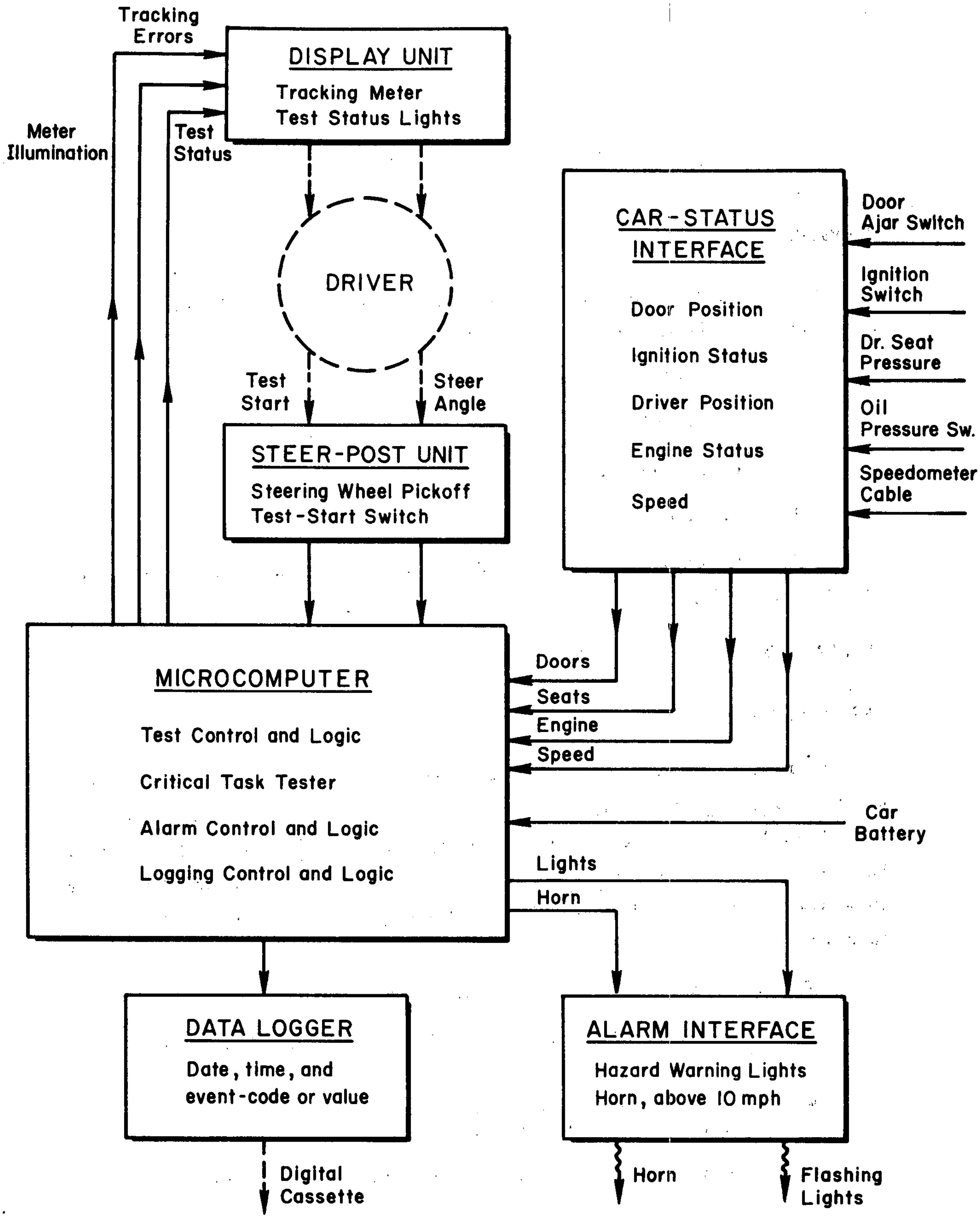


Figure F-1. Critical Task Interlock — Assemblies and Interfaces

- Reads the positions of experimenter-controlled switches which define certain parameters of the test
- Monitors the driver's position in the vehicle (to ensure he does not leave his seat)
- Monitors the vehicle's ignition switch, oil pressure switch (to determine whether the engine is running), and vehicle speed
- Activates the alarms (hazard lights and horn)
- Controls and sends data to the data logger
- Turns off system power 10 minutes after the ignition switch is turned off

The microcomputer functions are defined by a program stored on four programmable read-only memory (PROM) chips. The microcomputer is mechanized on a single 4-1/2 in. × 6 in. printed circuit card. The circuitry required to interface the microcomputer with the off-processor task electronics and data logger, and to drive the display meter, status lights, and meter illumination is mounted on a second printed circuit card.

Automobile Interface

Additional circuitry is required to interface the microprocessor with the following automobile sensors:

- Driver's seat pressure
- Door position
- Ignition switch
- Oil pressure
- Speed

This circuitry is mounted on a third printed circuit card, PC3, along with the horn and hazard warning light drivers. Power components of these circuits are mounted separately on a heat sink to enable removing excess heat from the instrumentation case in which the electronics are contained.

Power Supplies

System power is derived from the automobile battery. A DC to DC converter is used to generate ±15 VDC from the automobile 12 volt system. Two voltage regulators are used to derive +5 V and -10 V for the digital components and the heat sink on the back of the instrument case.

Instrument Case

All electronics, power supplies, and the data logger are contained in a locked instrument case mounted in the left rear of the automobile trunk. A heat sink is mounted on the lower backside of the case in a position which minimizes obstruction of free air flow around the heat sink.

The dimensions of the case are 18 in. L × 12 in. H × 12 in. D. This size was selected to fill the left rear volume of the Chevrolet Nova trunk. This configuration is designed to prevent storage of anything on top of or behind the case which would interfere with air circulation around the heat sink.

Front Panel Controls

Figure F-2 portrays the front panel of the CTI and data logger. Four switches provide the experimenter control over the test configuration and enable a self test feature of the system.

Two thumbwheel switches identified as λ PASS enable the experimenter to select the level of instability λ required to pass the test. λ PASS may be set in increments of 0.1 as high as 9.9.

The TEST STRATEGY thumbwheel switch enables the experimenter to choose the number of trials required to pass a given test and the number of trials in the test. A table is provided on the panel which lists the allowable combinations versus switch position.

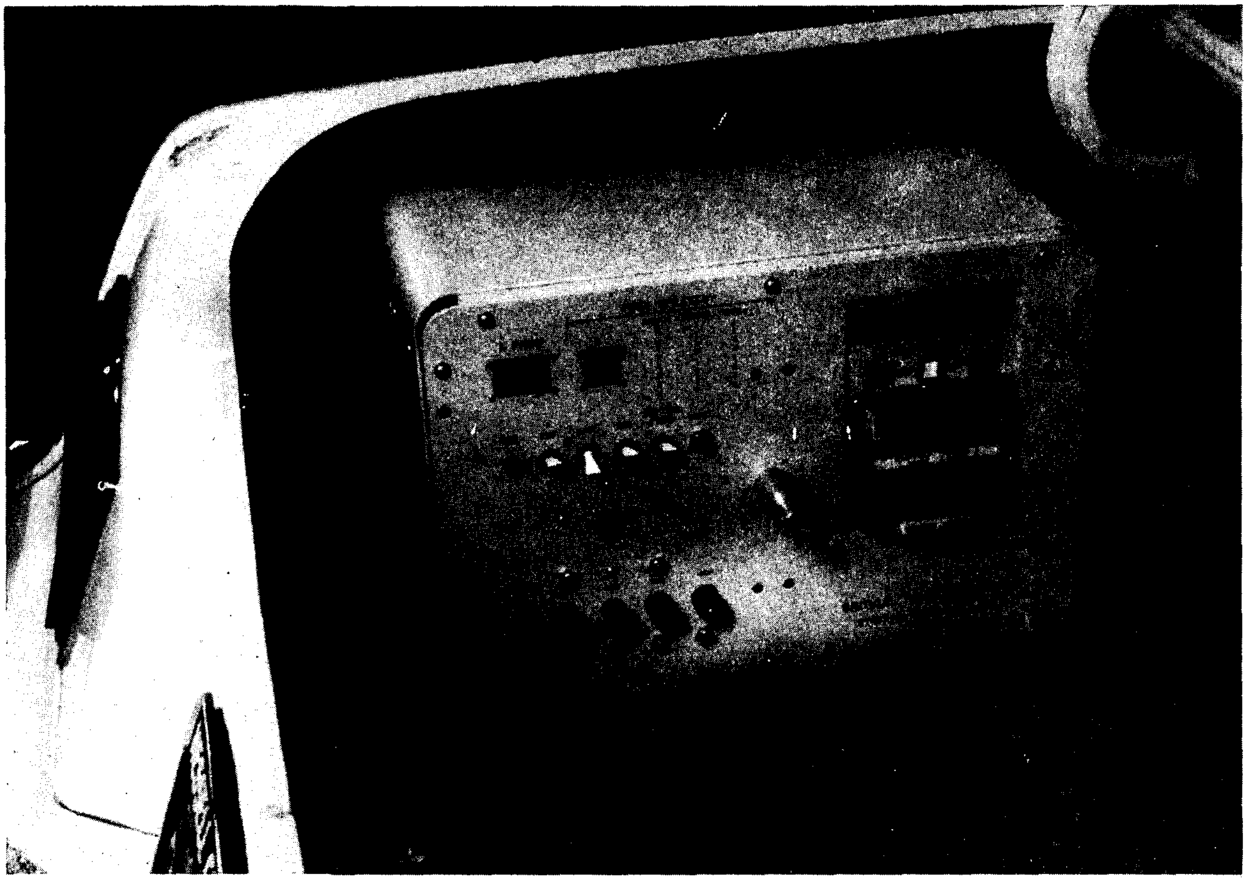


Figure F-2. Front Panel of Critical Task Interlock and Data Logger

The RUN/SELF-TEST toggle switch is put in the RUN position for normal operation. The tracking loop is closed internally when the RUN/TEST switch is put in the TEST position. In this mode, a trial will proceed automatically with the computer closing the loop until the trial is passed or instability reached, depending on the test settings. The system will score itself and proceed to the next trial. Various modes can be exercised by setting λ_{PASS} lower or higher than the self-test λ_c limit. This self-test feature should facilitate system checkout and troubleshooting.

The END TEST toggle switch enables the experimenter to select whether the test is terminated when the preset λ_{PASS} level of difficulty is reached (top position), or continues to the operator's "critical-instability limit" λ_c (bottom position).

The black switch below the row of white switches allows the alarms (i.e., horn and flashing lights) to be deactivated. Five fuses for ±15 V and 12 V complete the front panel configuration of the CTI.

On the right-hand side of the instrument case is a separate panel mounting the data logger and its controls. The clock controls are at the top of the panel. When the TIME SET switch is depressed, the experimenter initializes and starts the clock. The clock can be set to the nearest one-tenth of a minute.

A digital cassette stores the logged data. The cassette is loaded from the front panel below the clock controls and the TIME SET and LOAD controls.

CRITICAL TASK CONTROL/DISPLAY CONFIGURATION

The steering wheel pickoff assembly and the display are shown in Fig. F-3. The steering wheel potentiometer pickoff is geared to the steering wheel shaft in a 3:1 ratio. The assembly design is derived from a similar design which was developed by STI for other DOT test cars. The assembly is rugged and reliable, and does not interfere with the driver.

The Test Start momentary pushbutton switch is mounted on the left side of the steering wheel potentiometer assembly and is conveniently accessible to the driver.

The display consists of the display meter, scale, three status lights (TEST, PASS, and FAIL) and the meter illumination light. The specially compensated display meter frequency response is essentially flat from DC to approximately 10 Hz and is well damped. The display is enclosed in a case mounted to the steering wheel shaft housing as shown in Fig. F-3. This positioning minimizes interference of the driver's view of the dashboard. Placement of the meter and scale deep within the meter case localizes the field of view of the meter to the driver's head area and minimizes the possibility of a confederate trying to pass the test for the driver.

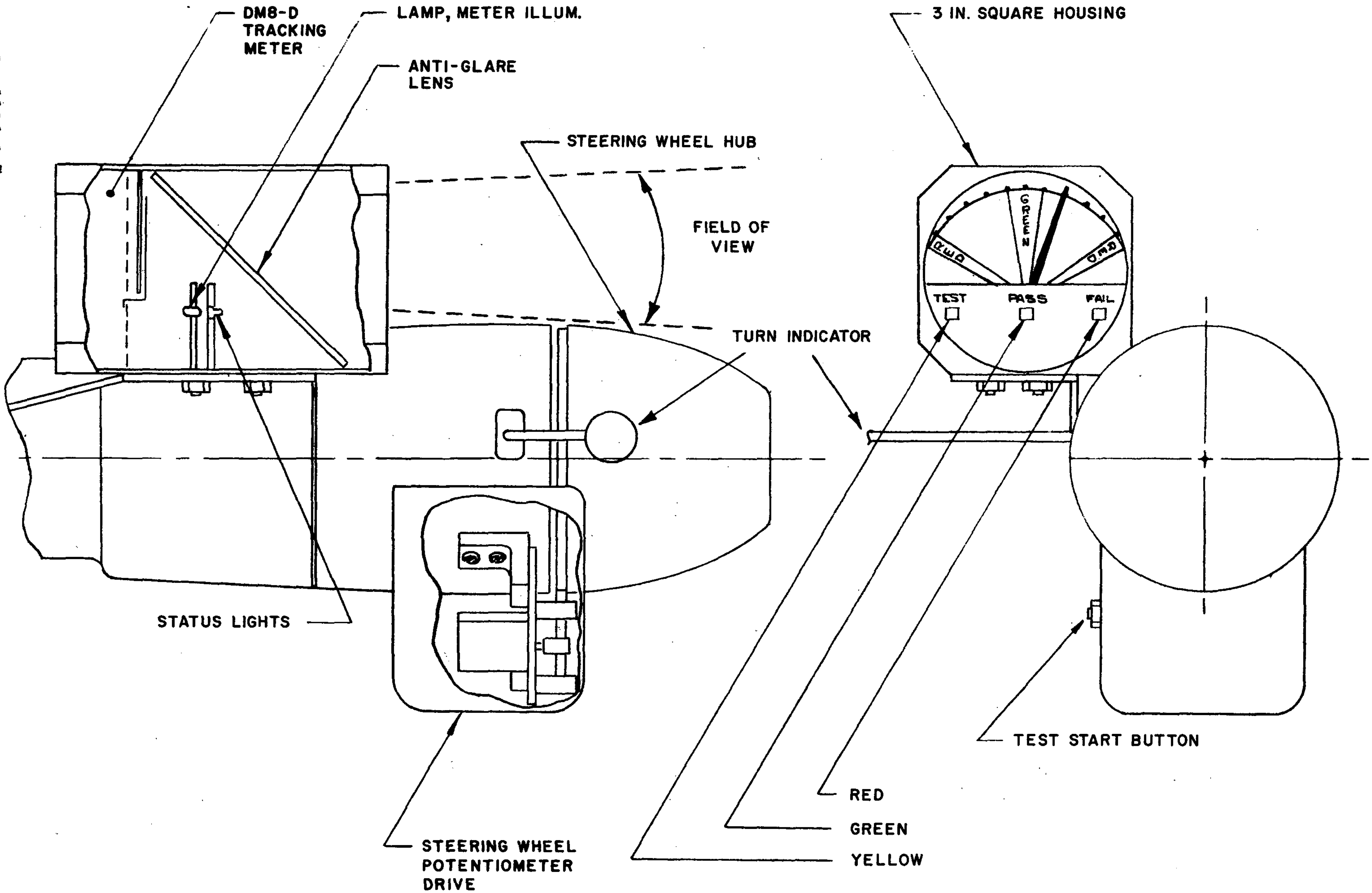


Figure F-3. Critical Task Interlock Display and Steering Wheel Pickoff

DATA LOGGER

The data logger records digital data and time of day upon external command. It is specifically designed for use in automotive vehicles and operates from the normal 12 volt vehicle power.

The data logger uses a digital tape cassette for the memory and is capable of storing in excess of 3000 data points. The recorder includes one Model 906 Formatter manufactured by Instrumentation Technology Corporation and one Model 201 Recorder manufactured by the Memodyne Corporation.

As can be seen in Fig. F-2, the time of day clock and day counter consists of two digits of day, two digits of hours, two digits of minutes, and one digit of tenths of a minute. A crystal controlled oscillator maintains the time and all updating is automatic. The initial time setting is accomplished by setting the correct time and day on the front panel thumbwheel switches and pressing the TIME SET button.

Data consist of two four-bit BCD numbers in sequence. Following the recording of data, the logger records the day count and the time of day.

The data sent to the recorder is coded according to the event which is to be recorded. Table F-1 shows the event code.

Given the Table F-1 data, several pieces of additional information can be inferred. When the CTT score is greater than the criterion, the driver has passed the test. Unusual combinations of events, such as test failure followed by a lengthy period before an ignition off event, may indicate inappropriate vehicle usage.

TABLE F-1. DDWS RECORDED AND COMPUTED DATA

a) Raw Data

Event	Data Logged
Ignition On Time, Day, Date	A, Strategy[a], Pass Criterion (λ_p),
Ignition Off	B, Time, Day, Date
Speed > 10 mph[b]	E, Time, Day, Date
End of Trial Time Day, Date	D, Weight (0, 1)[c], CTT Score (λ_c),

b) Typical Data Interpretation

$\lambda_c > \lambda_p$	=	Pass
A and B, no D	=	Did not take test
A, D, $\lambda_c < \lambda_p$, B, but no E	=	Failed test (practice or driving < 10 mph?)
A, D, $\lambda_c < \lambda_p$, B, and E	=	Driving with alarms on

[a]Set at 1 pass in 4 attempts in this study.

[b]When test not passed.

[c]Within (1) or outside (0) of preset band.