

**VALIDATION OF THE STANDARDIZED FIELD SOBRIETY TEST
BATTERY AT BACs BELOW 0.10 PERCENT**

FINAL REPORT

**Submitted to:
U.S. DEPARTMENT OF TRANSPORTATION
NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION**

**Jack Stuster
Marcelline Burns**

August 1998

**ANACAPA SCIENCES, INC.
P.O. Box 519
Santa Barbara, California 93102**

Technical Report Documentation Page

1. Report No. DOT HS ?	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle <i>Validation of the Standardized Field Sobriety Test Battery at BACs Below 0.10 Percent</i>		5. Report Date August 1998	
7. Author(s) Jack W. Stuster, PhD, CPE, and Marcelline Burns, PhD		6. Performing Organization Code: n/a 8. Performing Organization Report No. n/a	
9. Performing Organization Name and Address Anacapa Sciences, Inc. P.O. Box 519 Santa Barbara, CA 93102		10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTNH22-95-C-05192	
12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration 400 Seventh Street, SW Washington, D.C. 20590		13. Type of Report and Period Covered Final Report	
15. Supplemental Notes James F. Frank, PhD was the Contracting Officer's Technical Representative (COTR) for this project.		14. Sponsoring Agency Code	
<p>16. Abstract</p> <p>This study evaluated the accuracy of the Standardized Field Sobriety Test (SFST) Battery to assist officers in making arrest decisions for DWI at blood alcohol concentrations (BACs) below 0.10 percent. NHTSA's SFST battery was validated at 0.10 percent BAC in 1981. The trend to reduce statutory DWI limits to 0.08 percent BAC prompted this research project.</p> <p>The research was composed of several project tasks, including planning, site-selection, training, data entry, and data analysis, in addition to the actual conduct of a major field study. The City of San Diego, California, was selected as the site. Seven officers of the San Diego Police Department's alcohol enforcement unit were trained in the administration and modified scoring of NHTSA's SFST battery (i.e., Horizontal Gaze Nystagmus-HGN, Walk and Turn, and One Leg Stand). SFST scoring was adjusted: the observation of four HGN clues indicated a BAC 0.08 percent (rather than four clues indicating a BAC 0.10 percent), and the observation of two HGN clues indicated a BAC 0.04 percent. During routine patrols, the participating officers followed study procedures in administering SFSTs and completing a data collection form for each test administered. The officers' final step in each case was the administration of an evidentiary breath alcohol test.</p> <p>Data analysis found the SFSTs to be extremely accurate in discriminating between BACs above and below 0.08 percent. The mean estimated and measured BACs of the 297 motorists tested were 0.117 and 0.122, respectively; the difference between the means (0.005 percent BAC) is very small and operationally irrelevant. Further, analyses found the HGN test to be the most predictive of the three components of the SFST battery ($r=0.65$), however a higher correlation was obtained when the results of all three tests were combined ($r=0.69$).</p> <p>Decision analyses found that officers' estimates of whether a motorist's BAC was above or below 0.08 or 0.04 percent were extremely accurate. Estimates at the 0.08 level were accurate in 91 percent of the cases, or as high as 94 percent if explanations for some of the false positives are accepted. Officers' estimates of whether a motorist's BAC was above 0.04 percent but lower than 0.08 percent were accurate in 94 percent of the decisions to arrest and in 80 percent of cases overall. Also, the officers and prosecutors who were interviewed about the SFSTs found the test battery to be acceptable for field use to establish probable cause for DWI arrest.</p> <p>The results of this study provide clear evidence of the validity of the Standardized Field Sobriety Test Battery to discriminate at 0.08 percent BAC, using a slightly modified scoring procedure. Further, study results strongly suggest that the SFSTs also accurately discriminate at 0.04 percent BAC.</p>			
17. Key Words DWI, DUI, detection, field sobriety, performance tests, alcohol, law enforcement, highway safety		18. Distribution Statement	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages	22. Price

EXECUTIVE SUMMARY

This report documents the research activities and presents the results of a study conducted for the National Highway Traffic Safety Administration (NHTSA) to evaluate the accuracy of the Standardized Field Sobriety Test (SFST) Battery to assist officers in making arrest decisions and to discriminate blood alcohol concentrations (BACs) below 0.10 percent. NHTSA's SFST battery was validated at 0.10 percent BAC in 1981. The trend to reduce statutory DWI limits to 0.08 percent BAC prompted this research project.

DESCRIPTION OF THE RESEARCH

The research was composed of several project tasks, including planning, site-selection, training, data entry, and data analysis, in addition to the actual conduct of a major field study. The City of San Diego, California, was selected as the site of the field study. Seven officers of the San Diego Police Department's alcohol enforcement unit were trained in the administration and modified scoring of NHTSA's SFST battery (i.e., Horizontal Gaze Nystagmus, Walk and Turn, and One Leg Stand). SFST scoring was changed slightly: the observation of four horizontal gaze nystagmus (HGN) clues indicated a BAC 0.08 percent (rather than four clues indicating a BAC 0.10 percent), and the observation of two HGN clues indicated a BAC 0.04 percent. During routine patrols, the participating officers followed study procedures in administering SFSTs and completing a data collection form for each test administered during the study period. The officers' final step in each case was the administration of an evidentiary breath alcohol test.

RESULTS

The participating officers completed a total of 298 data collection forms; only one case was eliminated from analysis because the motorist refused all forms of BAC testing. Data analysis found the SFSTs to be extremely accurate in discriminating between BACs above and below 0.08 percent. The mean estimated and measured BACs of the 297 motorists tested were 0.117 and 0.122, respectively; the difference between the means (0.005 percent BAC) is very small and operationally irrelevant. Further, analyses found the HGN test to be the most predictive of the three components of the SFST battery ($r=0.65$), however a higher correlation was obtained when the results of all three tests were combined ($r=0.69$).

The results of decision analyses provide clear indication of SFST accuracy. Decision analyses found that officers' estimates of whether a motorist's BAC was above or below 0.08 or 0.04 percent were extremely accurate. Estimates at the 0.08 level were accurate in 91 percent of the cases, or as high as 94 percent if explanations for some of the false positives are accepted. Officers' estimates of whether a motorist's BAC was above 0.04 but under 0.08 were accurate in 94 percent of the decisions to arrest and in 80 percent of the relevant cases, overall.

Finally, the officers and prosecutors who were interviewed about the SFSTs found the test battery to be fully acceptable for field use to establish probable cause for DWI arrest.

IMPLICATIONS

The results of this study provide clear evidence of the validity of the Standardized Field Sobriety Test Battery to discriminate above or below 0.08 percent BAC, using a slightly modified scoring procedure. Further, study results strongly suggest that the SFSTs also accurately discriminate above or below 0.04 percent BAC.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
Background.....	1
THE RESEARCH	5
Task 1: Refined Work Plan.....	5
Task 2: Specified SFSTs and Revised Procedures.....	5
Task 3: Selected and Recruited Law Enforcement Agency.....	5
Task 4: Conducted the Field Validation Study.....	9
Tasks 5 and 6: Analyzed Data and Prepared Final Report.....	14
RESULTS	15
Evaluation of SFST Accuracy.....	15
Comparison of Means.....	15
Correlation Analyses.....	17
Decision Analyses.....	18
Evaluation of SFST Acceptability.....	23
IMPLICATIONS	25
How Accurately Do the Tests Discriminate?.....	25
Which of the Components of the SFST Battery Is/ Are the Best Predictors?.....	26
How Reliable, or Consistent, Are the Tests?.....	26
Are the Tests Usable By Officers? Are They Accepted by Officers and Prosecutors?.....	27
Conclusions.....	28
REFERENCES	29
APPENDIX A: STANDARDIZED FIELD SOBRIETY TESTING	31

LIST OF TABLES

Table		Page
1	COMPONENTS OF THE FIELD EXPERIMENT PLAN.....	9
2	ESTIMATED AND MEASURED BAC BY AGE CATEGORY	15
3	ESTIMATED AND MEASURED BAC BY GENDER.....	16
4	ESTIMATED AND MEASURED BAC BY DISPOSITION, AGE CATEGORY, AND GENDER	16
5	CORRELATIONS OF SFST SCORES TO ESTIMATED AND MEASURED BAC.....	17
6	SUMMARY OF FALSE POSITIVES.....	19
7	SUMMARY OF FALSE NEGATIVES.....	20

FIGURES

Figure		Page
1	Fatality rates per million miles traveled in the U.S.....	2
2	Sequence of major project tasks.....	5
3	Data collection form used in the validation study.....	12
4	Decision matrix at 0.08 percent BAC.....	18
5	Decision matrices at 0.08 percent BAC for each component test.....	21
6	Venn diagram of 188 cases 0.08 percent BAC.....	22
7	Decision matrix at 0.04 percent BAC.....	23

ACKNOWLEDGMENTS

The National Highway Traffic Safety Administration and Anacapa Sciences, Inc., are grateful to the managers and officers of the San Diego Police Department for their cooperation during the performance of this research project. Special thanks to Officer Doug English, project liaison, and Sergeant Bill Snyder, supervisor of the Traffic Division's alcohol-enforcement unit.

INTRODUCTION

Beginning in 1975, the National Highway Traffic Safety Administration (NHTSA) sponsored research that led to the development of standardized methods for police officers to use when evaluating motorists who are suspected of Driving While Impaired (DWI).¹ Beginning in 1981, law enforcement officers have used NHTSA's Standardized Field Sobriety Test (SFST) battery to help determine whether motorists who are suspected of DWI have blood alcohol concentrations (BACs) greater than 0.10 percent. Since that time, many states have implemented laws that define DWI at BACs *below* 0.10. This report presents the results of research performed to systematically evaluate the accuracy of NHTSA's SFST battery to discriminate above or below 0.08 percent and above or below 0.04 percent blood alcohol concentration.

The report is presented in four sections. This brief Introduction presents the objectives of the research, provides a summary of the relevant traffic safety issues, and discusses the historical context of the study. The second section of the report describes the research tasks that were performed. The third section presents the results of the study. The final section of the report discusses the implications of the study results.

BACKGROUND

Nearly 1.4 million people have died in traffic crashes in the United States since 1966, the year of the National Traffic and Motor Vehicle Safety Act (which led to the creation of NHTSA in 1970). During the late 1960s and early 1970s more than 50,000 people lost their lives each year on our nation's public roads; more than half of the motorists killed had been drinking. Traffic safety has improved considerably since that time: the annual death toll has declined to about 40,000, even though the numbers of drivers, vehicles, and miles driven all have greatly increased. The dramatic improvements in traffic safety are reflected in the change in fatality rate per 100 million vehicle miles traveled: The fatality rate fell from 5.5 in 1966 to 1.7 in 1996 (FARS--Fatal Analysis Reporting System--96), a 69 percent improvement. Figure 1 illustrates this important trend. When miles traveled are considered, the likelihood of being killed in traffic in 1966 was more than three times what it is today.

Despite the significant improvements in traffic safety during the past 17 years, an average of more than 115 people still die each day from motor vehicle crashes in the United States. It is estimated that 41 percent of the drivers who die in crashes have been drinking.

¹ Various terms are used throughout the United States for offenses involving drinking and driving. In this report, Driving While Impaired (DWI) is used to refer to all occurrences of driving at or above the legal blood alcohol concentration (BAC) limit of a jurisdiction.

An emphasis on DWI enforcement since 1980 has been a factor in the significant improvement in traffic safety, as represented by declining fatal and alcohol-involved crash rates. NHTSA-sponsored research contributed substantially to the improved condition, in part, by providing patrol officers with useful and scientifically valid information and training materials concerning the behaviors that are most predictive of impairment. In particular, NHTSA sponsored research that led to the development of a DWI detection guide that listed 20 driving cues and the probabilities that a driver exhibiting a cue would have a BAC of at least 0.10 percent (Harris et al., 1980; Harris, 1980). A similar study was conducted recently that identified 24 driving cues that are predictive of DWI at the 0.08 level (Stuster, 1997). NHTSA also sponsored research that led to the development of a motorcycle DWI detection guide (Stuster, 1993). NHTSA's DWI training materials, based on the results of these studies, have exposed the current generation of law enforcement officers in the U.S. to information critical to DWI enforcement by providing a systematic, scientifically valid, and defensible approach to on-the-road DWI detection.

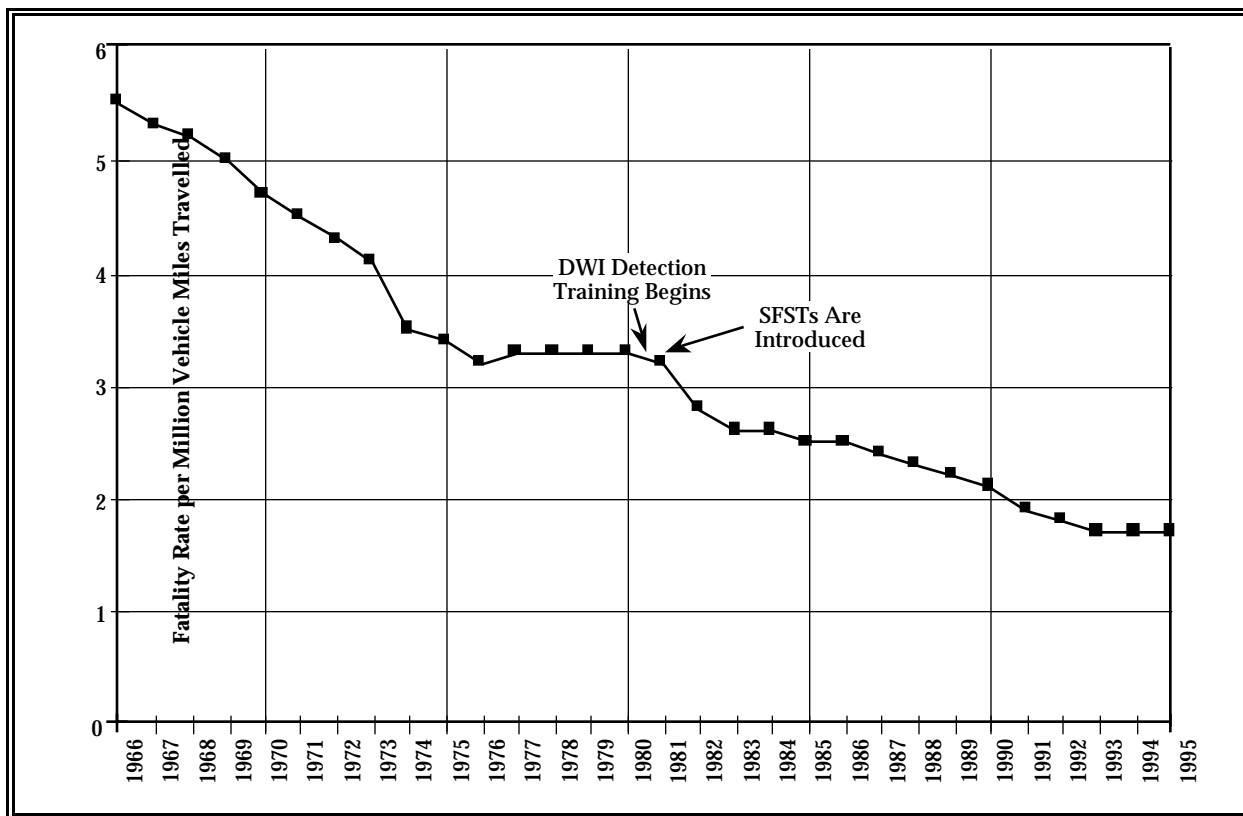


Figure 1. Fatality rates per million miles traveled in the U.S.

At the same time NHTSA was providing patrol officers with information concerning the driving behaviors that are the most predictive of impairment, the agency also sponsored research that led to the development of a standardized battery

of tests for officers to administer to assess driver impairment after an enforcement stop has been made. Drs. Marcelline Burns and Herbert Moskowitz conducted laboratory evaluations of several of the tests that were most frequently-used by law enforcement officers at the time (Burns and Moskowitz, 1977). In addition to a variety of customary roadside tests (e.g., finger-to-nose, maze tracing, backward counting), the researchers evaluated measures of an autonomic reaction to central nervous system depressants, known as horizontal gaze nystagmus. Horizontal gaze nystagmus (HGN) is an involuntary jerking of the eye that occurs naturally as the eyes gaze to the side. Aschan (1958) described studies that linked various forms of nystagmus to BAC, and Wilkinson, Kime, and Purnell (1974) reported consistent changes in horizontal gaze nystagmus with increasing doses of alcohol. At the time Burns and Moskowitz were conducting their seminal research for NHTSA, horizontal gaze nystagmus recently had been found to reliably predict BACs in a study conducted in Finland (Pentilla, Tenhu, and Kataja, 1974). Further, Lehti (1976) had just calculated a strong correlation between BAC and the onset of nystagmus.

All of the field sobriety tests evaluated by Burns and Moskowitz were found to be sensitive to BAC in varying degrees, at least under laboratory conditions. In addition, all of the tests showed a consistent increase in correlations with increasing BACs. Statistical analyses found the horizontal gaze nystagmus test to be the most predictive of the individual measures. However, the combined scores of three of the tests (One-Leg Stand, Walk-and-Turn, and Horizontal Gaze Nystagmus) provided a slightly higher correlation than the horizontal gaze nystagmus test by itself. The combined score correctly discriminated between BACs below or above 0.10 in 83 percent of the subjects tested in the original study (Burns and Moskowitz, 1977).

NHTSA immediately sponsored a subsequent study to standardize the test administration and scoring procedures and conduct further laboratory and field evaluations of the new battery of three tests. The researchers found that police officers tended to increase their arrest rates and were more effective in estimating the BACs of stopped drivers after they had been trained in the administration and scoring of the Standardized Field Sobriety Test battery. The results of this important study were documented in meticulous detail in the technical report, *Development and Field Test of Psychophysical Tests for DWI Arrest* (Tharp, Burns, and Moskowitz, 1981). That report has been cited throughout the U.S. to establish the scientific validity of the SFST battery and to support officers' testimony in court. NHTSA's SFST battery is described in Appendix A.

During the past 16 years, NHTSA's SFSTs largely have replaced the unvalidated performance tests of unknown merit that once were the patrol officer's only tools in helping to make post-stop DWI arrest decisions. Regional and local preferences for other performance tests still exist, even though some of the tests have not been validated. Despite regional differences in what tests are used to assist officers in making DWI arrest decisions, NHTSA's SFSTs presently are used in all 50 states. NHTSA's SFSTs have become the standard pre-arrest procedures for evaluating DWI in most law enforcement agencies.

The horizontal gaze nystagmus (HGN) test is considered by many law enforcement officers to be a foolproof technique (sometimes called a “silver bullet”) that provides indisputable evidence of alcohol in a motorist’s system. The normal variation in human physical and cognitive capabilities, and the effects of alcohol tolerance, result in uncertainties when arrest decisions are made exclusively on the basis of performance tests. These uncertainties have resulted in large proportions of DWI suspects being released rather than detained and transported to another location for evidentiary chemical testing. This is important because experienced drinkers often can perform physical and cognitive tests acceptably, with a BAC greater than 0.10 percent. However, most experienced drinkers cannot conceal the physiological effects of alcohol from an officer skilled in HGN administration. This is because horizontal gaze nystagmus is an involuntary reaction over which an individual has absolutely no control.

THE RESEARCH

This section provides a detailed description of all tasks performed during the field validation of the Standardized Field Sobriety Test Battery for use at 0.08 percent BAC. The technical approach to the research involved the performance of six major project tasks, as summarized in Figure 2 and described in the following pages.

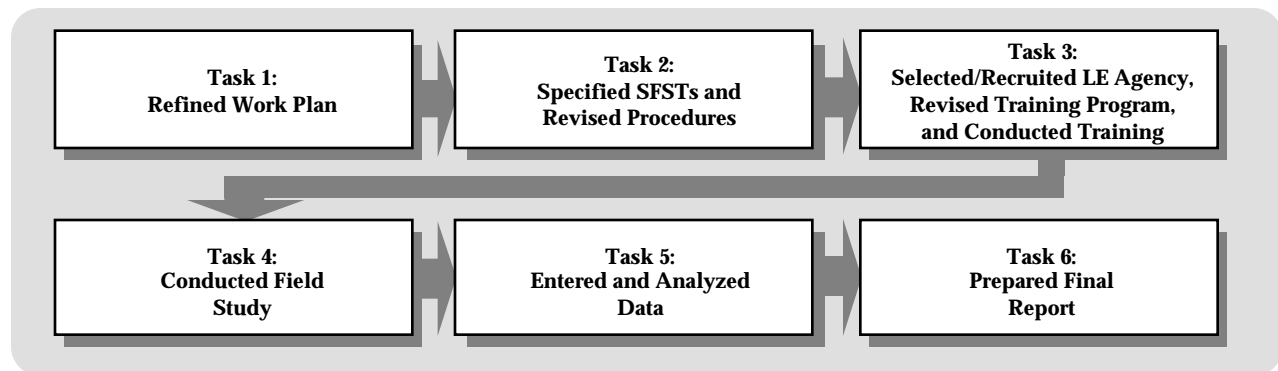


Figure 2. Sequence of major project tasks.

TASK 1: REFINED WORK PLAN

The objectives of the first project task were to meet with the Contracting Officer's Technical Representative (COTR) and other NHTSA SFST experts to discuss the project and to refine the proposed Work Plan based on those discussions. The project kick-off meeting was held at NHTSA headquarters on 24 October 1995. Substantive discussions with NHTSA personnel during and following the meeting contributed to the development of the technical approach described here.

TASK 2: SPECIFIED SFSTs AND REVISED PROCEDURES

Based on the widespread use and acceptance of NHTSA's Standardized Field Sobriety Test (SFST) Battery, validated at 0.10 percent BAC, NHTSA sponsored the current study to evaluate the SFSTs at lower BACs. The only modifications to be made to the SFSTs would be: 1) for officers to use the exhibition of four clues as an indication of BACs at the 0.08 level or greater (as officers presently are trained to use four clues as an indicator of BACs at 0.10 percent or greater), and 2) for officers to use the exhibition of two HGN clues as an indication of BACs greater than zero, but *below* 0.08 percent.

TASK 3: SELECTED AND RECRUITED LAW ENFORCEMENT AGENCY AND CONDUCTED TRAINING

This project task was composed of four subtasks, as described in the following paragraphs.

SUBTASK 3.1: IDENTIFIED SITE SELECTION CRITERIA

The site-selection criteria were:

- Candidate sites must employ lower legal BAC levels (0.08 for adults and zero tolerance for youth under 21 years).
- Candidate sites must generate a sufficient number of traffic enforcement stops and DWI arrests for accurate assessment of the tests' reliability and validity.
- Participating officers must have received NHTSA-approved SFST training from a certified instructor, possess at least one year of field experience administering SFSTs, and receive refresher training from project staff.
- Managers and officers of the participating law enforcement agency must agree to abide by the research procedures for the duration of the field study. For example, officers may use only the SFST Battery (and no other tests) together with their observations of the driver's general appearance and speech to make their arrest decisions; and, all test administrations must be recorded and submitted. Only agencies that could assure an extremely high level of cooperation and commitment would be recommended for participation.
- The site must have the capability of generating cases that represent the full range of alcohol experience. For example, a city with a disproportionate number of younger drivers might be more appropriate to ensure samples of sufficient size for the younger age categories.

SUBTASK 3.2: IDENTIFIED CANDIDATE SITES AND APPLIED SELECTION CRITERIA

Several factors constrained the site-selection process and limited the possible candidates for participation in this study. First, at the time the project was conducted, California, Oregon, and Utah were the only states that met both of the BAC-related site-selection criteria, namely a 0.08 BAC limit for DWI and a zero tolerance law for drivers under 21 years of age. Second, it was important to restrict the data collection period, to the extent possible, because it was believed that an extremely long data collection period might result in officers deviating from the study procedures. Strict adherence to study procedures was considered essential to ensuring the internal validity of the study.

The site-selection strategy adopted was to recruit a police department that serves one large city--a city large enough to generate a sufficient number of SFST administrations for statistical analysis by itself. A large city also was likely to have a traffic division with a dedicated DWI unit composed of trained experts. Focusing on traffic enforcement specialists would permit us to restrict participation in the study to officers who already had received NHTSA-approved SFST training and had additional field experience administering the test battery. Prior training in SFST administration was an important site-selection and methodological issue.

In the study that validated the SFST battery in 1981, all officers of an agency could participate, following training provided by the researchers. The procedure followed during the original study was appropriate then because no other officers (anywhere) had yet to receive the training. However, that procedure could not be followed in the current study because thousands of officers have received SFST

training since 1981. Only trained and experienced test administrators could be permitted to participate in the *current* study to avoid confounding study results with the effects of substantially different officer skill and experience levels in SFST administration and scoring. Officers who are formally trained and experienced in SFST administration tend to be concentrated in traffic enforcement and special DWI units.

This site-selection strategy was judged to provide the best approach to achieve the objectives of the current study, and the City of San Diego, California, was identified as the leading candidate community when the site-selection criteria were applied. The San Diego Police Department serves a resident population of more than one million, with a much larger service population attributable to tourism and several local military installations. The manner in which the San Diego Police Department satisfied the site-selection criteria is outlined below.

Number of SFST Administrations

The San Diego Police Department maintains a traffic division composed of 50 officers, including ten officers and a sergeant who form the alcohol enforcement unit. The alcohol enforcement unit deploys four or five officers on each night, Wednesday through Sunday. The time necessary to complete the associated paperwork usually limits each officer to a maximum of two DWI arrests each night. This results in about 130 arrests by officers of the special unit during a four week period. The other members of the traffic division, combined, make an additional 130 DWI arrests each month. San Diego Police Department officers do not hesitate to arrest drivers for BACs below 0.08 percent if they exhibit any evidence of impairment, even though low-BAC arrests usually are not prosecuted by the local district attorney.

Demographic Considerations

The Work Plan discussed the importance of selecting a site that offers cases for analysis that represent the full range of driver ages and BACs of interest. It was believed that a younger, rather than an older, driver population would result in more cases of zero tolerance violations and more SFST administrations overall. In this regard, San Diego and the surrounding area is home to four major US Navy bases and both the Navy and Marine Corps training centers. The area also is home to three major universities and several smaller colleges and technical schools.

Willingness to Participate

Naturally, formal approval by senior managers is required before any law enforcement agency can participate in a traffic safety study. Further, a manager's personal interest in a study that results in command emphasis concerning participation greatly contributes to the success of a project because of the quasi-military organizational structure of law enforcement agencies. That is, if managers believe participation to be of value to an agency they will direct their officers to follow the study procedures. In this regard, the commanding officer and other senior managers of the San Diego Police Department expressed their considerable interest in the study and directed their personnel to cooperate with the study team.

Command emphasis is an important component to ensure adherence to study procedures, but it is not sufficient; the participating officers also must be committed to the study. The willingness of a law enforcement agency to participate in a traffic safety study also can be measured, although subjectively, by the attitudes of field officers when discussing the general and specific issues involved in the study. The officers of the San Diego Police Department with whom we spoke about the field validation expressed genuine interest in the study and eagerness to be selected for participation.

Finally, the requirement for an agency to modify its established procedures to accommodate special study procedures usually is somewhat negotiable in a traffic safety study, but deviations from established study procedures were not negotiable in this field validation. It was explained that police managers and all participating officers must agree to abide by the study procedures to ensure the internal validity of study results. This was an area for concern to the project team because the San Diego Police Department's established DWI procedures included administering three field sobriety tests in addition to the three NHTSA SFSTs. A firm study requirement was that no other tests be administered to subjects because they might influence an officer's BAC estimates; that is, all officer-estimates of BAC must be based *exclusively* on results of the NHTSA SFST battery using the slightly modified scoring system. In this regard, San Diego police managers inquired with their district attorney and DWI supervisors, those who might object to the restriction, and found no opposition. In fact, it was mentioned that restricting sobriety testing to the three SFSTs would help streamline the procedures for everyone.

Prior SFST Training

All members of the San Diego Police Department's special alcohol-enforcement unit previously had received SFST training that was administered according to NHTSA-approved procedures and curriculum by certified DWI instructors. Although approximately half of the other members of the Traffic Division also had received SFST training, it was determined that the alcohol-enforcement unit would generate a sufficient number of SFST administrations for statistical analysis. All of the participating officers would receive a four-hour refresher training course prior to beginning the field study.

SUBTASK 3.3: RECRUITED LAW ENFORCEMENT AGENCY TO PARTICIPATE IN THE STUDY

NHTSA reviewed the site recommendations and approved San Diego as the site for the field study. Further discussions were held with managers and officers of the San Diego Police Department and a Memorandum of Agreement was signed that specified all study procedures and requirements.

SUBTASK 3.4: DEVELOPED SFST TRAINING PROGRAM

The experimental requirement that all participating officers be both trained and experienced in SFST administration eliminated the need to develop a special training program for this study. It was considered essential that the existing, NHTSA-approved SFST training program remain the training standard for the field evaluation. Because all participating officers already had received NHTSA-approved

SFST training, only a refresher program would be required. A four-hour refresher-training program was developed, based on the (October 1995) NHTSA curriculum. The purposes of the refresher training were to instruct the officers concerning the modified scoring system and obtain confirmation that all participants were administering and scoring the SFST battery correctly before beginning the field study.

TASK 4: CONDUCTED THE FIELD VALIDATION STUDY

Systematic evaluation of the SFSTs to assist officers in making arrest decisions at BACs below 0.10 percent, under field conditions, was the ultimate objective of this research. Although existing tests were the subject of the evaluation, the reasons for conducting the field study were the same as if the tests previously had not been validated. First, it was necessary to determine the accuracy of the modifications to test scoring, compared to actual BAC levels measured through other means. For cases in which the driver was arrested for DWI, correspondence would be assessed between scored performance on the SFSTs and BAC, as determined by breath test (blood and urine tests were discouraged but used if subjects refused to comply with breath testing). For cases in which a subject was administered SFSTs but then released on the basis of low estimated BAC, hand-held breath testing devices were used to establish actual BAC. The second purpose of the evaluation was to identify problems with test application in the field, which might include test administration, scoring procedures, or other factors that might affect the use of the tests by law enforcement personnel. Third, the courts' acceptance of evidence gathered using the slightly revised scoring procedures in the field evaluation would be assessed.

SUBTASK 4.1: PREPARED FIELD EXPERIMENT PLAN

A Field Experiment Plan was developed and approved by NHTSA to guide the conduct of the field study. The plan included the seven components depicted in Table 1 and discussed below.

TABLE 1
COMPONENTS OF THE FIELD EXPERIMENT PLAN

Component 1: Subjects
Component 2: Independent Variables
Component 3: Criterion Measures
Component 4: Materials
Component 5: Procedures
Component 6: Controls
Component 7: Data Analyses

Components 1 and 2: Subjects and Independent Variables

The primary independent variable of interest, BAC, was inextricably linked to the subjects in this study. Specifically, the experiment plan focused on obtaining data from adult motorists who were suspected of exceeding the legal limit of 0.08 percent BAC and youths under 21 who were suspected of exceeding the "zero-tolerance"

legal limit of 0.00. The accuracy of the SFSTs to discriminate at 0.08 and 0.04 percent BAC could not be assessed without data from individuals who had BACs over and *under* these values. Therefore, it was important to obtain BAC estimates from individuals who had both passed and failed the standardized field sobriety tests.

Component 3: Criterion Measures

The only appropriate criterion measure to assess the accuracy of SFSTs is BAC. Measures of impairment are irrelevant because performance of the SFSTs must be correlated with BAC level, rather than driving performance. BAC provides an objective and reliable measure that states have recognized as presumptive and/or per se evidence of impairment, depending on the statute. To obtain these criterion measures, it was determined that all drivers who were administered the SFST Battery must be tested for BAC, regardless of the results of the SFSTs. In other words, it would be essential to test the individuals who were judged to have BACs below the relevant statutory level and who subsequently would be released. Participating officers were instructed concerning the importance of obtaining BAC data for all subjects, in order to calculate the accuracy of the tests.

All police officers participating in the study were equipped with NHTSA-approved, portable breath testing devices to assess the BACs of all drivers who were administered the SFSTs, including those who were released without arrest. Further, arrested subjects were tested both in the field with a portable device and at the booking site. The use of passive alcohol sensors (PAS) during the study was not permitted.

Component 4: Materials

Only the existing SFSTs were to be administered, which require no equipment. A pen, pencil, or small flash light frequently are used by officers as a stimulus or target for the HGN test, but a finger can be used with equal effectiveness.

The data collection form used in the study is presented as Figure 3. The data collection form was extremely important in this study for several reasons. As is the case in most field studies, the form must be as simple to complete as possible to minimize the workload of participating officers. In the present case, it also was important for the form to be designed to guide the officer in the administration of the SFSTs, to facilitate standardization and systematic scoring of the tests. In addition, the form designed for this study had to both encourage and provide assurances that officers had followed the study procedures. Most important, it was essential that officers would conduct a breath test and record actual subject BAC as the final step of the process; that is, actual BACs were to be entered on the form only after BAC estimates based on SFST performance had been recorded. Hand-held breath testing devices with digital displays were used for this purpose.

Component 5: Procedures

The sixth component of the field experiment plan was the specification of procedures to be used for administering the tests and obtaining independent measures of BAC. The procedures to be followed by participating officers were listed

as a series of six numbered steps on the data collection form that was used in the field study. The study procedures were to be followed whenever a participating officer suspected an adult driver of being alcohol impaired or a youth under 21 of having a BAC greater than zero. In practice, officers administered the SFSTs to all motorists who exhibited *any* objective behavior or other cue associated with having consumed alcohol, even if impairment was not evident. A breath, blood, or urine test was administered to all motorists who performed the SFSTs, but only after the officer had made an arrest/no arrest decision based on the officer's scoring of the driver's SFST performance, and recorded a BAC estimate. The data collection form structured the procedure by presenting all officer actions as a series of numbered steps. Requiring officers to record the time of BAC estimates and BAC tests ensured that officers' estimates were not influenced by the results of the chemical tests. Completed data collection forms were sent to Anacapa Sciences on a weekly basis for data entry.

In some states, such as California, officers have the right to administer a breath test to a driver who has exhibited any objective sign of alcohol-consumption. Compliance is mandatory if the officer can articulate a reasonable suspicion of the motorist having consumed alcohol (such as the odor of an alcoholic beverage). SFSTs were administered only to drivers who exhibited some objective DWI cue, thus, no problems were experienced in obtaining BAC data, even from subjects whose SFST performance was acceptable. The field breath test was conducted as the final step after the SFST procedure was completed, which is the *de facto* procedure followed by most officers who are equipped with field breath testing devices.

To further ensure compliance with study procedures, the participating law enforcement officers signed a statement affirming that they would abide by the established study procedures. In addition, project staff monitored the data collection effort, periodically riding along with participating officers to ensure that study procedures were being followed.

Component 6: Controls

Extraneous variables that could affect the outcome of the study must be controlled to the extent possible. The controls that were implemented to ensure the validity of study results have been discussed in this section, including systematic procedures and the use of only trained and experienced officers.

Component 7: Data Analyses

The data analysis plan was designed to answer the following research questions.

- How accurately do the tests discriminate between subjects who are above or below 0.08 and 0.04 percent BACs?
- Which of the components of the SFST battery is/are the best predictor(s) of BAC?
- How reliable, or consistent, are the tests?
- Are the tests usable by police officers? Are they readily accepted by officers and prosecutors?

NHTSA/ANACAPA SFST VALIDATION DATA FORM

Officer ID: _____ Driver: Adult Male
 Under 21 Female
↳ Age: _____

Month _____ Day _____ 1996 Time of Stop: _____ hr _____ min

FIELD SOBRIETY TESTS ADMINISTERED

1. HORIZONTAL GAZE NYSTAGMUS TEST

Clues

	Right Eye	Left Eye
Lack of smooth pursuit	<input type="checkbox"/>	<input type="checkbox"/>
Nystagmus at maximum deviation	<input type="checkbox"/>	<input type="checkbox"/>
Nystagmus onset before 45 degrees	<input type="checkbox"/>	<input type="checkbox"/>
Clues	<input type="checkbox"/>	<input type="checkbox"/>

+ =

Total HGN Clues (6 clues maximum) ←

4 or more 0.08 / 2 or more 0.04

2. ONE LEG STAND TEST (seconds)

	0-10	11-20	21-30
Sways while balancing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uses arms for balance	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hops to maintain balance	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Puts foot down	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cannot perform test (4 clues -- maximum)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Total One Leg Stand Clues	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2 or more 0.08

3. WALK AND TURN TEST

Clues

Loses balance while listening to instructions	<input type="checkbox"/>
Starts before instructions are finished	<input type="checkbox"/>
	1st 9 2nd 9
Stops while walking	<input type="checkbox"/>
Does not touch heel to toe	<input type="checkbox"/>
Steps off the line	<input type="checkbox"/>
Raises arms for balance	<input type="checkbox"/>
Incorrect number of steps	<input type="checkbox"/>
Trouble with turn (explain) _____	<input type="checkbox"/>
Cannot perform the test (8 clues -- maximum)	<input type="checkbox"/>
Total Walk and Turn Clues	<input type="checkbox"/>

2 or more 0.08

4. ESTIMATE OF BAC BASED ON SFSTs:

Time of estimation _____ hr _____ min

5. SUBJECT BAC Refused

PBT → Time of PBT test _____ hr _____ min

Other → Time of other test _____ hr _____ min

↳ Breath Blood Urine

6. DISPOSITION: Warning Citation DUI Arrest

Figure 3. Data collection form used in the validation study.

SUBTASK 4.2. TRAINED OFFICERS IN THE USE OF THE SFSTs

Dr. Marcelline Burns, one of the investigators who developed the SFST battery, developed and conducted the refresher training for the participating officers. Dr. Burns' research and training experience in this field ensured that officers received effective and credible refresher instruction. Dr. Burns was assisted in the training session by the project director and NHTSA COTR.

SUBTASK 4.3. IMPLEMENTED EXPERIMENTAL DESIGN AND COLLECTED DATA

Implementation of the experiment design began immediately following the completion of officer refresher training on 23 May 1996 and continued through 9 November. Specific study procedures were:

- Only officers who were members of the San Diego Police Department's alcohol-enforcement unit and who received NHTSA-approved SFST training participated directly in the study. Dr. Marcelline Burns provided brief "refresher" training to all participating officers to ensure a consistent and systematic approach to SFST administration during the study.
- Upon commencement of the study period, participating officers used only the SFST Battery (i.e., Horizontal Gaze Nystagmus, Walk and Turn, One Leg Stand) together with their observations of a driver's general appearance and speech, to establish inferences about a subject for whom there was reasonable suspicion of driving while impaired. In other words, no tests other than the three SFSTs were performed.
- Participating officers performed the administration steps in the sequence specified on the data collection form; that is, they,
 1. Administered the **Horizontal Gaze Nystagmus** test and recorded results.
 2. Administered the **One Leg Stand** test and recorded results.
 3. Administered the **Walk and Turn** test and recorded results.
 4. Used the scoring systems that were printed on the data collection form (by counting test "clues") to estimate the subject's BAC. Recorded their estimate of the subject's BAC based on SFST performance, together with their observations of the subject's general appearance and speech. Also, they recorded the time when their estimate was made.
 5. Checked the box that indicated the disposition of the stop: Warning, Citation, or Arrest.
 6. Recorded the subject's BAC obtained from a field breath test; or, checked the appropriate box for other tests or responses. Blood and urine test results were provided later; every effort was made to obtain a breath test result for all subjects. Recorded the time when the BAC test was performed.
- Obtained a BAC for all subjects who were administered SFSTs as the final step in the test administration procedure. BACs were obtained for all subjects tested including those subjects who officers estimated, on the basis of SFST results, to have BACs below the legal limit.
- Participating officers completed and submitted a data collection form for each subject tested during the study period; that is, all administrations of the SFST battery by

participating officers were recorded on a data collection form and submitted for analysis.

- All completed data collection forms were sent to Anacapa Sciences, Inc., for data entry and analysis.

SUBTASK 4.4 CONDUCTED COURT AND POLICE INTERVIEWS

The final data collection task was the conduct of open-ended interviews with participating police officers and prosecutors who were exposed to the new SFSTs during DWI cases. The purposes of the interviews were to determine if the tests were acceptable to the officers for use in the field and to the prosecutors for use of test results in court.

TASKS 5 AND 6: ANALYZED DATA AND PREPARED FINAL REPORT

All data collection forms were returned to Anacapa Sciences, Inc., sequentially numbered, and the contents entered into a computerized data base. Data analyses were performed by the project director and Dr. Marcelline Burns. The results of those analyses are presented in the following section of this report.

RESULTS

This study was conducted to evaluate the accuracy of NHTSA's Standardized Field Sobriety Test Battery in assisting officers to make arrest decisions at BACs above and below 0.08 percent under field conditions. A secondary objective of the study was to evaluate the possibility that the test battery also could be used to assist officers in making arrest decisions at BACs lower than 0.08 percent.

The seven participating officers from the San Diego Police Department's alcohol-enforcement unit completed a total of 298 data collection forms during the study period; only one case was eliminated from analysis because the subject refused to submit to any form of BAC testing. Officer compliance with study procedures and motivation to participate in the study remained high throughout the data collection period.

EVALUATION OF SFST ACCURACY

Three methods were used to evaluate the accuracy of the SFST battery to discriminate at the BACs of interest: comparison of means, correlation analyses, and decision analyses.

COMPARISON OF MEANS

Table 2 presents a summary of the estimated and measured BAC data by age category. The table shows that 91.9 percent of the motorists tested were adults, compared to 8.1 percent youth, defined as motorists under the age of 21 years. The mean estimated and measured BACs of the younger motorists were approximately 0.035 lower than the BACs of the adults tested during the field study. The officers' mean estimated BACs, however, were very close to the mean measured BACs for both adults and youth; on average, the difference between officers' estimates and the actual BACs were only 0.005 percent for adults and 0.007 percent for youth.

TABLE 2
ESTIMATED AND MEASURED BAC (%) BY AGE CATEGORY

Age Category	Number	Percent	Estimated BAC (Mean)	Measured BAC (Mean)
Adults	273	91.9	0.120	0.125
Youth	24	8.1	0.083	0.090
Total	297	100.0	0.117	0.122

Table 3 presents a summary of the estimated and measured BAC data by gender category. The table shows that 87.9 percent of the motorists tested were males, compared to 12.1 percent females, with adults and youth combined. The mean estimated BACs of the male and female motorists tested were identical (i.e., 0.117 percent). Again, for both categories, the officers' mean estimated BACs were very close to the mean measured BACs; on average, the difference between officers' estimates and the actual BACs were only 0.004 percent for males and 0.012 percent for females.

TABLE 3
ESTIMATED AND MEASURED BAC (%) BY GENDER

Gender	Number	Percent	Estimated BAC (Mean)	Measured BAC (Mean)
Male	261	87.9	0.117	0.121
Female	36	12.1	0.117	0.129
Total	297	100.0	0.117	0.122

Table 4 presents a more detailed accounting of the estimated and measured BAC data by age and gender category, and by the disposition of the enforcement stop. In addition, the table shows that 73 percent of all motorists who were tested during the field study were arrested for DWI based on SFST performance and officer evaluations. Approximately 22 percent of the motorists tested received warnings and five percent were cited for a motor vehicle violation other than DWI.

TABLE 4
ESTIMATED AND MEASURED BAC (%) BY DISPOSITION, AGE CATEGORY, AND GENDER

Disposition & Category	Number	Percent	Estimated BAC (Mean)	Measured BAC (Mean)
Warnings	65	21.9	0.060	0.044
Adults	57		0.063	0.045
Male Adults	53		0.063	0.044
Female Adults	4		0.070	0.054
Youth	8		0.036	0.038
Male Youth	6		0.037	0.038
Female Youth	2		0.035	0.040
Citations	15	5.1	0.055	0.046
Adults	11		0.050	0.040
Male Adults	9		0.047	0.043
Female Adults	2		0.065	0.029
Youth	4		0.070	0.062
Male Youth	2		0.060	0.055
Female Youth	2		0.080	0.070
Arrests	217	73.0	0.138	0.150
Adults	205		0.139	0.152
Male Adults	180		0.139	0.150
Female Adults	25		0.139	0.160
Youth	12		0.119	0.135
Male Youth	11		0.121	0.134
Female Youth	1		0.100	0.140
Total	297	100.0	0.117	0.122

The data presented in Table 4 also show that officers tended to slightly over-estimate the BACs of motorists who had lower BACs, and slightly under-estimate BACs at the higher levels. Overall, however, officers' estimates were extremely accurate. Based on SFST results and officers' observations, the officers' mean estimated BAC of the 297 motorists was 0.117 percent, compared to the mean measured BAC of 0.122. Although statistically significant, the difference of 0.005 percent BAC is a trivial and operationally irrelevant under-estimate of actual BACs that is within the margin of error of sophisticated evidentiary testing equipment.

CORRELATION ANALYSES

The accuracy of the SFSTs was further evaluated by conducting a series of correlation analyses to identify the degree to which officers' individual estimates of BAC corresponded with subjects' actual, or measured, BAC. A correlation coefficient is a statistic, usually represented as r , that expresses the relatedness of two variables, that is, the degree to which the variables co-vary. In this case, the two variables were an officer's estimate and the subject's actual BAC. The Pearson product-moment correlation method was used to calculate the relationship between these variables; cases with complete SFST results ($n=261$) were used in this analysis.

If officers had predicted the precise BACs of all subjects (to three decimal points), the correlation coefficient would be +1.00; the correlation coefficient would be zero if there were no relationship between the estimated and actual BACs. For predictive measures, especially those administered under field conditions, a correlation of 0.65 to 0.70 is considered to be very high.

Table 5 presents the results of the correlation analyses. The table shows that HGN test results had the highest correlation with measured BAC of the three components of the SFST battery ($r=0.65$). However, a slightly higher correlation was obtained when the results of the three component tests were combined ($r=0.69$). The table also shows strong correlations between test results and officers' estimated BACs, indicating that officers were following procedures and interpreting test results correctly. All of the correlations were found to be statistically significant ($p=.005$).

TABLE 5
CORRELATIONS OF SFST SCORES TO ESTIMATED AND MEASURED BAC (%)
N=261 CASES WITH COMPLETE SFST SCORES

Rank	SFST(s)	Correlation (r) with Estimated BAC	Correlation (r) with Measured BAC
1	3 Tests Combined	0.75	0.69
2	HGN	0.71	0.65
3	Walk-and-Turn	0.64	0.61
4	One Leg Stand	0.61	0.45

DECISION ANALYSES

The third method used to evaluate the accuracy of the SFST battery was to construct a decision matrix that describes the four possible combinations of the two variables of interest, estimated and actual BACs above and below the levels of interest. Figure 4 presents the first decision matrix, with the four major cells of the matrix representing the four possible decisions at 0.08 percent BAC. The numbers in the major cells are the number of cases for each type of decision out of the 297 SFST administrations. The two shaded cells represent correct decisions based on SFST results: 1) 210 motorists who officers estimated to have BACs equal to or greater than 0.08 percent, who later were found to have BACs \geq 0.08 by BAC testing (by breath, blood, or urine analysis); and, 2) 59 motorists who officers estimated to have BACs below 0.08 percent, who later tested below 0.08.

Figure 4 also reveals the incorrect decisions: 1) 24 motorists who officers estimated to have BACs greater than 0.08 who later were found to have BACs below that level (false positives); and, 2) four subjects who officers estimated to have BACs below 0.08 who later tested above 0.08 (false negatives).

It can be calculated from the data contained in Figure 4 that officers' decisions were accurate in 91 percent of the 297 cases (i.e., $[210+59] \div 297 = .906$). Further, officers' decisions to arrest were correct in 90 percent of the cases in which BAC was estimated to be \geq 0.08 (i.e., $210 \div 234 = .897$), and decisions not to arrest were correct in 94 percent of the cases in which BAC was estimated to be below 0.08 (i.e., $59 \div 63 = .937$). These results indicate a high degree of accuracy, but it will be instructive to consider more closely those cases in which incorrect decisions were made.

		Officers' Estimated BACs		
		<0.08%	0.08%	
Measured BACs	0.08%	n=4	n=210	n=214
	<0.08%	n=59	n=24	n=83
		n=63	n=234	N=297

Accurate in 91% of cases overall
90% accurate in "yes" decisions
94% accurate in "no" decisions

Figure 4. Decision matrix at 0.08 percent BAC.

Table 6 presents a summary of the data for each of the 24 false positives (FPs). These cases are labeled False Positives because the officers estimated the subjects' BACs to be 0.08 percent, but subsequent testing found BACs below 0.08. However, in several cases, officers were correct in identifying impairment, which probably influenced their estimates of BAC.

TABLE 6
 SUMMARY OF FALSE POSITIVES

	Case Number	Estimated BAC (%)	Number of HGN Clues	Measured BAC (%)	Is Estimate Consistent with Clues?
1	30	0.08	4	0.050	yes
2	34	0.08	4	0.058	yes
3	121	0.08	6	0.060	yes
4	186	0.08	4	0.063	yes
5	226	0.08	6	0.058	yes
6	227	0.08	4	0.060	yes
7	129	0.09	4	0.070	yes
8	175	0.09	4	0.070	yes
9	32	0.09	6	0.076	yes
10	127	0.09	6	0.028	yes
11	224	0.10	4	0.070	yes
12	16	0.10	6	0.070	yes
13	196	0.10	6	0.074	yes
14	52	0.11	4	0.050	yes
15	178	0.12	6	0.070	yes
16	246	0.12	6	0.069	yes
17	12	0.08	2	0.060	no
18	164	0.08	2	0.070	no
19	165	0.08	2	0.020	no
20	135	0.08	3	0.078	no
21	137	0.09	n/a	0.030	?
22	75	0.09	2	0.048	no
23	104	0.09	3	0.037	no
24	13	0.12	0	0.043	no

In 16 of the cases listed in Table 6, the officers' estimates of BAC were consistent with the number of HGN clues observed (i.e., four or more HGN clues to

support an estimate 0.08), however, the motorists subsequently were found to have actual BACs below 0.08 percent. In seven of the cases, the officers' estimated BACs were inconsistent with the number of HGN clues observed. It is important to note that six of the 24 false positives had measured BACs of 0.07 percent, and three had BACs greater than 0.07 but less than 0.08 (i.e., 0.074, 0.076, and 0.078). All nine of these BACs are within the margin of error of the testing devices. Further, Case Number 16 was a juvenile (0.069), which rendered the difference between estimated and measured BACs irrelevant in a zero tolerance jurisdiction; that is, it was a correct arrest decision despite the BAC estimate. In addition, two of the subjects with measured BACs of 0.07 were arrested for DWI, because the officers' believed that they were too impaired to be permitted to drive. Finally, Case Number 30, with an estimated BAC of 0.08 and a measured BAC of 0.05 percent, was found to be a psychiatric patient, which helped to explain her erratic behavior, poor SFST performance, and apparent impairment.

Although the proportions of correct decisions presented in Figure 4 reflect a high degree of accuracy, the accuracy of officers' decisions is even better if some of the borderline cases are accepted. An accuracy rate of 94 percent for all officer decisions based on SFST results was calculated by including as correct decisions Case 16 (the youth with a 0.069 percent BAC) and the nine false positives with BACs between 0.07 and 0.08, discussed in the previous paragraph.

Table 7 summarizes the four cases in which officers estimated the subjects' BACs to be below 0.08 percent, but later found the measured BACs to be 0.08. Six HGN clues would be expected for Case Number 193 (0.10 percent) and Case Number 99 (0.12 percent). It is unknown why the officers observed only two HGN clues. In contrast, officers recorded four HGN clues for Case Number 131 and Case Number 114, which would indicate BACs greater than 0.08, however, the officers' estimated-BACs were only 0.06 percent. It is unknown why the officers did not follow the test interpretation guidelines in these two cases; their low estimates probably reflect other observations made in combination with SFST performance.

TABLE 7
SUMMARY OF FALSE NEGATIVES

	Case Number	Estimated BAC (%)	Number of HGN Clues	Measured BAC (%)	Is Estimate Consistent with Clues?
1	193	0.06	2	0.100	yes
2	99	0.06	2	0.120	yes
3	131	0.06	4	0.080	no
4	114	0.06	4	0.116	no

Similarly, in seven of the false positive cases listed previously in Table 6, officers apparently did not follow the test interpretation guidelines; that is, fewer than four HGN clues were reported, yet the officers' estimated-BACs were at least

0.08 percent. It is possible that other factors influenced the officers' estimates. For example, the subjects might have appeared to be more impaired than indicated by HGN results as a consequence of prescription or recreational drugs taken in addition to alcohol.

A series of decision analyses was performed to calculate the contributions of the component tests of the battery to officers' estimates of BAC. Figure 5 presents three decision matrices, one for each of the SFSTs. The matrices are similar to the one in Figure 4, but with the criterion numbers of clues at 0.08 percent BAC substituted for officers' estimates. Figure 5 shows the HGN test to be the most accurate independent predictor of whether a motorist's BAC is above or below 0.08 percent.

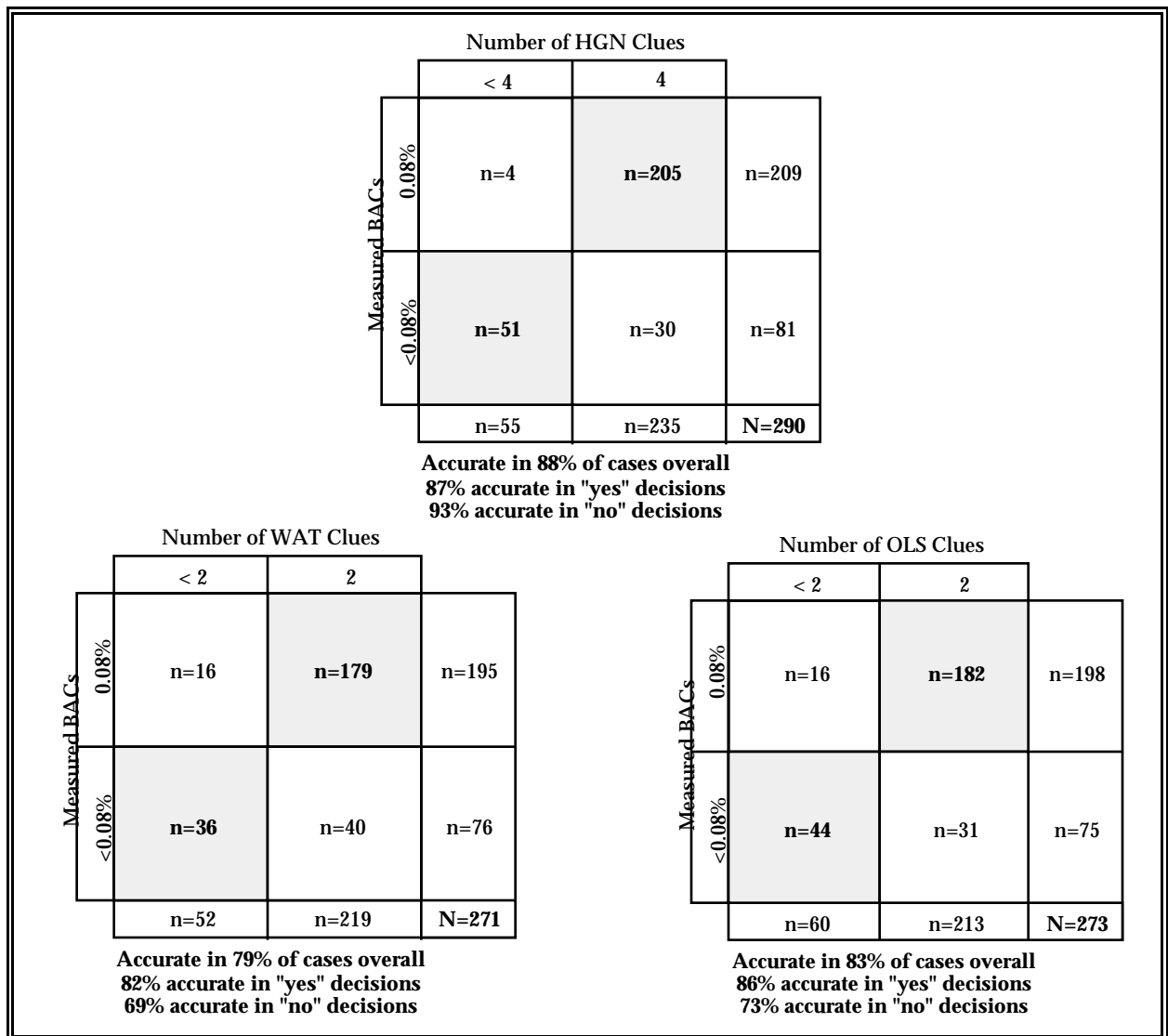


Figure 5. Decision matrices at 0.08 percent BAC for each component test of the SFST battery.

Further analyses were performed to explore methods for combining the results of the three component tests. Only the 261 cases that included test results for all three component tests could be used in this analysis. Of those cases, 73 were found to have BACs below 0.08 percent and 188 cases had measured BACs 0.08 percent. In 162 of the 188 cases (86 percent), all three component SFSTs were unanimous in their predictions.

Figure 6 presents a Venn diagram that illustrates the contributions of the three tests to the 14 percent of cases in which a discrepancy occurred. The figure shows there were 162 cases with BACs 0.08 in which all three SFSTs indicated a BAC 0.08 (the number outside the circles in Figure 6), and 26 cases in which one or more test disagreed (the numbers inside the circles). A single test indicated a BAC below 0.08 in 17 of the cases (8+2+7), and two tests were involved in nine of the cases (1+1+7). There were no cases in which all three tests predicted incorrectly.

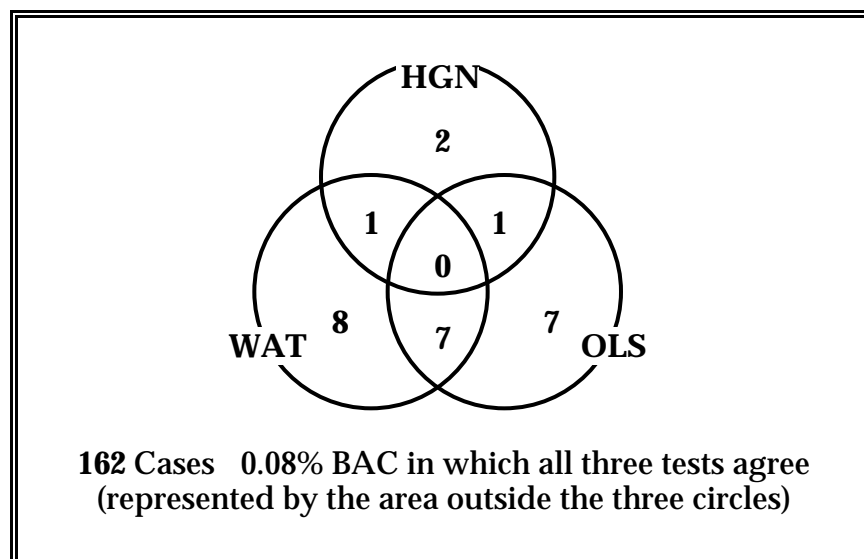


Figure 6. Venn diagram of 188 cases 0.08% BAC; 26 cases in which all three tests do not agree.

The horizontal gaze nystagmus test (HGN in the diagram) was about four times less likely to be the source of a discrepancy than the other two tests. Only two of the single-test discrepancies were attributable to HGN results, compared to eight cases for the Walk and Turn test (WAT), and seven cases for the One Leg Stand (OLS). Overall, the HGN test was involved in only four of the discrepancies, compared to 16 cases for the Walk and Turn and 15 cases for the One Leg Stand.

The question of the SFST battery's accuracy in discriminating BACs above and below 0.04 percent is addressed by the following decision matrix, presented in Figure 7; the shaded cells of the matrix again represent correct decisions based on SFST results. The figure shows that officers estimated motorists' BACs to be equal to or greater than 0.04 but under 0.08 percent in 54 cases, and in 51 of those cases their estimates were found to be correct by subsequent breath, blood, or urine testing;

these values result in an accuracy rate of 94 percent for these decisions (i.e., $51 \div 54 = .94$). The figure also shows that officers estimated that 29 motorists had BACs below 0.04, and in 15 of those cases their estimates were found to be correct by subsequent testing, resulting in a 52 percent accuracy rate ($15 \div 29 = .52$). Overall, officers were accurate in 80 percent of the cases when discriminating between subjects who were above 0.04 but below 0.08 percent BAC (i.e., $[51 + 15] \div 83 = .80$).

		Officers' Estimated BACs		
		<0.04%	0.04 <0.08%	
Measured BACs	0.04 <0.08%	n=14	n=51	n=65
	<0.04%	n=15	n=3	n=18
		n=29	n=54	N=83

Accurate in 80% of cases overall
94% accurate in "yes" decisions
52% accurate in "no" decisions

Figure 7. Decision matrix at 0.04 percent BAC.

EVALUATION OF SFST ACCEPTABILITY

In interviews and during ride-along observations, the officers who participated in the study fully accepted the SFSTs for evaluating motorists for DWI at BACs below 0.10 percent. All of the officers were formally trained in SFST administration and scoring and all had sufficient field experience to develop confidence in their abilities to discriminate at the 0.08 level. Further, it was the officers' experience with the SFST battery that the component tests could be administered to all but a small proportion of drivers and under all reasonable environmental conditions.

Interviews also were conducted with representatives of the San Diego City Attorney's Office to inquire concerning the acceptability of the SFSTs to prosecutors and judges in DWI cases. The attorneys interviewed reported that none of the 298 DWI arrests made by participating officers during the study period was negatively affected by the SFST battery, or by excluding the other tests that traditionally had been used by the department.

The attorneys further explained that as prosecutors they normally prefer as much evidence as possible, and in a DWI case more tests usually generate more evidence they can use. However, it has been their recent experience that a test used by another local law enforcement agency has negatively affected cases they have prosecuted. Defense attorneys have been unsuccessful in their challenges of NHTSA's SFST battery, but they have successfully challenged the validity of the other test because it has not been evaluated in a systematic and scientific manner. Prosecutors who were interviewed suggested that the optimum situation would be for all law enforcement agencies to restrict their field sobriety evaluations to the same standardized battery of three tests.

IMPLICATIONS

The research documented in this report found that NHTSA's Standardized Field Sobriety Test Battery accurately and reliably assists officers in making DWI arrest decisions at 0.08 percent BAC. The study also found that the SFSTs can be used to assist officers in making arrest decisions at 0.04 percent BAC by using two HGN clues as the criterion rather than four clues, which is the criterion for a 0.08 percent or above BAC determination. The primary implication of the study results is that the SFST battery is a valid method for making roadside DWI decisions at 0.08 and 0.04 percent BAC. Specific implications of the study results are presented in the following paragraphs in response to the research questions listed previously.

HOW ACCURATELY DO THE TESTS DISCRIMINATE BETWEEN SUBJECTS WHO ARE ABOVE OR BELOW 0.08 AND 0.04 PERCENT BACs?

This study found NHTSA's SFST battery to be an accurate method for discriminating motorists' BACs above and below 0.08 percent and above and below 0.04 percent, when the tests are conducted by trained officers, as summarized below.

COMPARISON OF MEANS

The mean estimated BAC of the 297 motorists included in the study was 0.117 percent, compared to the mean measured BAC of 0.122. The difference of 0.005 percent BAC (i.e., five one-thousandths of a percent BAC) is very small and operationally irrelevant. The accuracy of officers' estimates during this study, in large measure, confirms the anecdotal accounts and observations of officers in the field that suggest remarkable abilities to predict a motorists' BAC on the basis of SFST results.

CORRELATION ANALYSES

Correlation analyses found the HGN test to be very predictive of measured BACs ($r=0.65$). A higher correlation was obtained when the results of the three component tests were combined ($r=0.69$). All of the correlations are statistically significant, meaningful, and in the rank order expected from previous SFST research.

DECISION ANALYSES

Decision analyses found that officers' estimates of whether a motorist's BAC was above or below 0.08 or 0.04 percent were extremely accurate. Estimates at or above the 0.08 level were accurate in 91 percent of the cases, or as high as 94 percent if explanations for ten of the false positives are accepted. Estimates at or above the 0.04 level (but below 0.08) were accurate in 94 percent of the relevant cases. It is important to note that officers' decisions not to arrest were more accurate at 0.08 than at 0.04 (94 percent compared to 52 percent).

Although the relatively small number of low BACs in the data base ($n=83$) might constrain confidence in the SFSTs at the 0.04 level, the data strongly suggest

operational utility to accurately discriminate above or below 0.04 percent BAC. Further, these results are consistent with the results of a recent study conducted to evaluate the SFST battery for use by officers in Colorado.

Colorado has a two-tier statute that permits officers to arrest motorists for driving under the influence (DUI) if found to have a BAC \geq 0.10 percent, and for a lesser offense, driving while ability impaired (DWAI), if found to have a BAC \geq 0.05 but below 0.099 percent. Of the 234 drivers tested during the Colorado study for whom BACs were known, 93 percent of the officers' decisions to arrest at the 0.05 percent criterion were correct, and 64 percent of the decisions to release were correct. Overall in the Colorado study, 86 percent of the officers' decisions at the 0.05 level were correct, based on SFST results (Burns and Anderson, 1995; Anderson and Burns, 1997).

WHICH OF THE COMPONENTS OF THE SFST BATTERY IS/ARE THE BEST PREDICTOR(S) OF BAC?

The horizontal gaze nystagmus test was found to be the most predictive of the three component tests, but correlations with measured BACs were higher when the results of all three tests were combined, as reported earlier. The implications of this study result are that all components of the SFST battery should be administered when possible or practical. However, the data indicate that the HGN test alone can provide valid indications to support officers' arrest decisions at both 0.08 and 0.04 percent BAC.

HOW RELIABLE, OR CONSISTENT, ARE THE TESTS?

Reliability is a measurement concept that represents the consistency with which a test measures a type of performance or behavior. In the current context, a reliable field sobriety test provides consistent results when administered to the same individual by two different officers, under nearly identical conditions. This type of "inter-rater" reliability was impossible to measure directly during this study, due to the constraints imposed by field conditions. In particular, it would have been unrealistic to subject motorists to the SFST battery twice, or to require that officers operate in pairs during their patrols.

Evidence of SFST reliability can be found in the results of the previous laboratory studies, in which the constraints on repeated measure were eliminated by the use of paid subjects and officers. Tharp, Burns, and Moskowitz (1981) found relatively high inter-rater reliability for BAC estimates based on SFST results ($r=.72$). The researchers also found that inter-rater reliability increased in subsequent sessions ($r=.80$), indicating the important role of training and experience in achieving accuracy, reliability, and overall proficiency.

In addition, correlation coefficients, in general, are measures of reliability. For this reason, the correlations between estimated and actual BACs obtained during the field study ($r=.69$) indicate a high degree of reliability for tests designed to be administered at roadside.

ARE THE TESTS USABLE BY POLICE OFFICERS UNDER A VARIETY OF ROADSIDE CONDITIONS? ARE THEY READILY ACCEPTED BY OFFICERS AND PROSECUTORS?

All of the officers who participated in this study were members of the San Diego Police Department's alcohol enforcement unit, all had previously received NHTSA-approved training in DWI detection and SFST administration, and all had at least three years of experience in the Traffic Division before joining the special unit. Prior to beginning the field study, the officers demonstrated competence in the administration of the component tests and interpretation of test results. Participation was limited to members of the alcohol-enforcement unit of a single law enforcement agency. These experience and training requirements were imposed, to control variables, to the extent possible, that might affect study results.

As a consequence of the selection criteria, all participating officers were proficient in the use of the SFST battery. The officers reported that they use their SFST skills daily in their work, and their experience has made them confident in the ability of the test battery to discriminate at 0.08 percent BAC, and at lower levels. Further, officers reported that the tests can be administered in all reasonable environmental conditions. In short, the officers who participated in this study consider the SFST battery to be extremely useful, in fact, essential tools for the performance of their professional duties.

The prosecutors interviewed during the study reported that the SFST battery has been acceptable to them and the courts because it was developed and validated in a systematic and scientific manner. They suggested that all law enforcement agencies should limit officers to use of the SFST battery in performance evaluations of DWI because other tests usually lack credibility in court. No problems were experienced in any of the 298 cases resulting from the field study, indicating the SFSTs to be fully acceptable to the courts in establishing probable cause to arrest a motorist for DWI.

NOTE ABOUT THE ACCEPTABILITY OF THE HGN TEST

Many law enforcement officers from across the United States have reported their sincere appreciation to NHTSA for developing the SFST battery, and in particular, the horizontal gaze nystagmus test. However, some officers have expressed frustration about the resistance of some courts to accept HGN results, despite the clear and unequivocal support of scientific research and field experience. It is likely that this remaining resistance to the horizontal gaze nystagmus test is attributable to a misunderstanding concerning the purpose of a field sobriety test, and can be explained by reference to "face validity," a term used in the behavioral sciences to describe one component of a measure's acceptability.

Many individuals, including some judges, believe that the purpose of a field sobriety test is to measure driving impairment. For this reason, they tend to expect tests to possess "face validity," that is, tests that appear to be related to actual driving tasks. Tests of physical and cognitive abilities, such as balance, reaction time, and information processing, have face validity, to varying degrees, based on the

involvement of these abilities in driving tasks; that is, the tests seem to be relevant “on the face of it.” Horizontal gaze nystagmus lacks face validity because it does not appear to be linked to the requirements of driving a motor vehicle. The reasoning is correct, but it is based on the incorrect assumption that field sobriety tests are designed to measure driving impairment.

Driving a motor vehicle is a very complex activity that involves a wide variety of tasks and operator capabilities. It is unlikely that complex human performance, such as that required to safely drive an automobile, can be measured at roadside. The constraints imposed by roadside testing conditions were recognized by the developers of NHTSA’s SFST battery. As a consequence, they pursued the development of tests that would provide statistically valid and reliable indications of a driver’s BAC, rather than indications of driving impairment. The link between BAC and driving impairment is a separate issue, involving entirely different research methods. Those methods have found driving to be impaired at BACs as low as 0.02 percent, with a sharp increase in impairment at about 0.07 percent (Moskowitz and Robinson, 1988; Stuster, 1997). Thus, SFST results help officers to make accurate DWI arrest decisions even though SFSTs do not *directly* measure driving impairment.

Horizontal gaze nystagmus is the most accurate diagnostic of BAC available to officers in the field. HGN’s apparent lack of face validity to driving tasks is irrelevant because the objective of the test is to discriminate between drivers above and below the statutory BAC limit, not to measure driving impairment. Throughout the United States, DWI laws permit arrest decisions to be made on the basis of the statutory BAC limit, irrespective of a specific motorist’s degree of impairment. Motorists also can be arrested at BACs *below* the statutory limit if their driving performance is demonstrably impaired by alcohol or other drugs.

CONCLUSIONS

The results of this study provide clear evidence of the validity of the Standardized Field Sobriety Test Battery to discriminate above or below 0.08 percent BAC. Further, study results strongly suggest that the SFSTs also accurately discriminate above or below 0.04 percent BAC.

Finally, in addition to establishing the validity of the SFST battery, this study has found the tests to be acceptable, indeed welcomed, by law enforcement officers and DWI prosecutors.

REFERENCES

- Anderson, E.W. and Burns, M. (1997). Standardized Field Sobriety Tests: A Field Study. *Proceedings of the 14th International Conference on Alcohol, Drugs and Traffic Safety Volume 2*, 635-639.
- Aschan, G. (1958). Different types of alcohol nystagmus. *Acta Otolaryngology, Supplement 140*, 69-78.
- Burns, M. and Anderson, E.W. (1995). A Colorado Validation Study of the Standardized Field Sobriety Test (SFST) Battery. Colorado Department of Transportation.
- Burns, M. and Moskowitz, H. (1977). *Psychophysical Tests for DWI Arrest*. U.S. Department of Transportation, National Highway Traffic Safety Administration Final Report DOT-HS-5-01242, Washington, D.C.
- Harris, D.H., Dick, R.A., Casey, S.M., and Jarosz, C.J. (1980). *The Visual Detection of Driving While Intoxicated*. U.S. Department of Transportation, National Highway Traffic Safety Administration Final Report DOT-HS-7-1538.
- Harris, D.H. (1980). Visual detection of driving while intoxicated. *Human Factors*, 22(6), 725-732.
- Lehti, H.M.J. (1976). The effects of blood alcohol concentration on the onset of gaze nystagmus. *Blutalkohol*, Vol. 13, 411-414.
- Moskowitz, H., and Robinson, C.D. (1988). *Effects of Low Doses of Alcohol on Driving-Related Skills: A Review of the Evidence*. U.S. Department of Transportation, National Highway Traffic Safety Administration, DOT-HS-807-280, Washington, D.C.
- Pentilla, A., Tenhu, M., and Kataja, M. (1971). *Clinical Examination For Intoxication In Cases of Suspected Drunken Driving*. Statistical and Research Bureau of TALJA. Iso Roobertinkatu 20, Helsinki, Finland.
- Stuster, J.W. (1997). *The Detection of DWI at BACs Below 0.10*. U.S. Department of Transportation, National Highway Traffic Safety Administration Final Report (in press).
- Stuster, J.W. (1993). *The Detection of DWI Motorcyclists*. U.S. Department of Transportation, National Highway Traffic Safety Administration Final Report DOT-HS-807-839.

Tharp, V., Burns, M., and Moskowitz, H. (1981). *Development and Field Test of Psychophysical Tests for DWI Arrest*. U.S. Department of Transportation, National Highway Traffic Safety Administration Final Report DOT-HS-805-864, Washington, D.C.

Wilkinson, I.M.S., Kime, R., and Purnell, M. (1974). Alcohol and human eye movement. *Brain*, 97, 785-792.

APPENDIX A
STANDARDIZED FIELD SOBRIETY TESTING

STANDARDIZED FIELD SOBRIETY TESTING

The Standardized Field Sobriety Test (SFST) is a battery of three tests administered and evaluated in a standardized manner to obtain validated indicators of impairment and establish probable cause for arrest. These tests were developed as a result of research sponsored by the National Highway Traffic Safety Administration (NHTSA) and conducted by the Southern California Research Institute. A formal program of training was developed and is available through NHTSA to help police officers become more skillful at detecting DWI suspects, describing the behavior of these suspects, and presenting effective testimony in court. Formal administration and accreditation of the program is provided through the International Association of Chiefs of Police (IACP). The three tests of the SFST are:

- Horizontal gaze nystagmus (HGN),
- Walk-and-turn, and
- One-leg stand.

These tests are administered systematically and are evaluated according to measured responses of the suspect.

HGN TESTING

Horizontal gaze nystagmus is an involuntary jerking of the eye which occurs naturally as the eyes gaze to the side. Under normal circumstances, nystagmus occurs when the eyes are rotated at high peripheral angles. However, when a person is impaired by alcohol, nystagmus is exaggerated and may occur at lesser angles. An alcohol-impaired person will also often have difficulty smoothly tracking a moving object. In the HGN test, the officer observes the eyes of a suspect as the suspect follows a slowly moving object such as a pen or small flashlight, horizontally with his or her eyes. The examiner looks for three indicators of impairment in each eye: if the eye cannot follow a moving object smoothly, if jerking is distinct when the eye is at maximum deviation, and if the angle of onset of jerking is within 45 degrees of center. If, between the two eyes, four or more clues appear, the suspect likely has a BAC of 0.10 or greater. NHTSA research indicates that this test allows proper classification of approximately 77 percent of suspects. HGN may also indicate consumption of seizure medications, phencyclidine, a variety of inhalants, barbiturates, and other depressants.

WALK AND TURN

The walk-and-turn test and one-leg stand test are “divided attention” tests that are easily performed by most unimpaired people. They require a suspect to listen to and follow instructions while performing simple physical movements. Impaired persons have difficulty with tasks requiring their attention to be divided between simple mental and physical exercises.

In the walk-and-turn test, the subject is directed to take nine steps, heel-to-toe, along a straight line. After taking the steps, the suspect must turn on one foot and

return in the same manner in the opposite direction. The examiner looks for eight indicators of impairment: if the suspect cannot keep balance while listening to the instructions, begins before the instructions are finished, stops while walking to regain balance, does not touch heel-to-toe, steps off the line, uses arms to balance, makes an improper turn, or takes an incorrect number of steps. NHTSA research indicates that 68 percent of individuals who exhibit two or more indicators in the performance of the test will have a BAC of 0.10 or greater.

ONE LEG STAND

In the one-leg stand test, the suspect is instructed to stand with one foot approximately six inches off the ground and count aloud by thousands (One thousand-one, one thousand-two, etc.) until told to put the foot down. The officer times the subject for 30 seconds. The officer looks for four indicators of impairment, including swaying while balancing, using arms to balance, hopping to maintain balance, and putting the foot down. NHTSA research indicates that 65 percent of individuals who exhibit two or more such indicators in the performance of the test will have a BAC of 0.10 or greater.

COMBINED MEASURES

NHTSA's SFST training materials instruct officers in the use of the following decision table for combining the results of the HGN and Walk and Turn test.

		HGN Clues							
		0	1	2	3	4	5	6	
Walk and Turn Clues	0								
	1								
	2								
	3								
	4								
	5								
	6								
	7								
	8								
	9								

Along the top of the table, circle the number of the subject's HGN clues. Along the left side of the table, circle the number of the subject's Walk and Turn clues. Draw a line down from the number of HGN clues and a line across from the number of Walk and Turn clues. If the intersection is within the shaded area, the subject has a BAC 0.10 percent.