



DWI

By Phillip B. Price, Sr. & Spurgeon Cole, Ph.D.

NHTSA Field Sobriety Tests Validation v. Invalidation

Dr. Marcelline Burns and cohorts recently conducted and published three new studies on the Standardized Field Sobriety Tests (SFST). These studies were funded in whole or in part by the National Highway Traffic Safety Administration (NHTSA). As indicated, NHTSA titled each of the reports as validation studies of the SFST. Here, we will give a short description of how each study was conducted, and then we will discuss the requirements of studies and tests from an analytical point of view to determine if the studies were conducted in a proper manner. We will point out some interesting facts about each study followed by our conclusion.

'A Colorado Validation Study Of the Standardized Field Sobriety Test (SFST) Battery'

The Colorado "study" was submitted in November 1995 by Marcelline Burns, Ph. D., and Colorado Deputy Ellen W. Anderson. It was funded by NHTSA and conducted in and around Aspen, Colorado (Pitkin County Sheriff's Office). Officers administered the SFST (HGN, W&T, OLS) and made an arrest decision for DUI based upon their belief that the subject had a blood alcohol content (BAC) of .05 percent or higher. Allegedly, no preliminary breath tests (PBT) were used, however, observers were present only 41 percent of the time.² Observers had a PBT. The law enforcement agencies volunteered for the five-month study. A total of 305 subjects were tested. The authors claim that overall the officers were correct 86 percent of the time in their decision to arrest or release. According to the authors' technical summary, the Colorado study demonstrates that "the SFSTs are valid tests, i.e., they serve as indices of the presence of alcohol at impairing levels."³

They make no claim that the SFST can predict driving ability. In fact, Dr. Burns has specifically stated that the tests are not able to detect driving impairment.⁴

In the Colorado study, 79 percent of the subjects had BACs of .05 percent or above.⁵ The base rate or "guess rate" then, is 79 percent. Base rate or guess rate refers to the number or percentage of subjects in the study that had a BAC above .05 percent. When the base rate is

79 percent, the effectiveness of the test is measured in terms of incremental validity. The employment of the SFST as well as other observations only improved decision accuracy by 7 percent. In other words, the officers back at the station having no access to the SFST scores or other observations, would come within seven percentage points of making as many correct decisions as the officers in the field. Since 79 percent of the subjects had a BAC above .05, an officer can simply arrest everyone in the sample and be correct 79 percent of the time. It is of particular interest to note that the mean BAC in the Colorado study was .152 percent.⁶ Over half had a BAC of .15 percent or higher and 25 percent of the individuals had a BAC of .20 or higher.⁷ These individuals should be considered "locks" and consequently this spuriously inflates the reported accuracy of the study. After all, only those drivers that were already thought to be impaired in the study were included in the study by the officers.⁸ It is also noteworthy that in the Colorado study the officers used a criterion in the walking phase of the walk and turn test that is not part of the NHTSA scoring procedure. The study indicates that "improper balance" was the criterion most often observed. It is not one of the clues listed in the NHTSA approved scoring manuals. The walk and turn was only 61 percent reliable as an indicator of BAC level in the Colorado study. Unfortunately, the researchers report no validity or reliability scores for the SFST.

A small gem is buried within the study. It is stated that age and "nervousness in the circumstances of potential arrest" can affect balance and coordination.⁹ This nervousness is indeed a very significant factor to reckon with in the SFST battery.¹⁰ Only 18 percent of the subjects were female and the subject pool consisted of no females between the ages of 51 and 70 years. The most disturbing finding is the fact that the



PHILLIP B. PRICE, SR. concentrates on DWI/DUI defense in Huntsville, Alabama. He is a fellow with the National College for DUI Defense and has also served as Dean and Regent to the college. He is a Past President to the Alabama Criminal Defense Lawyers Association (ACDLA). The author of many articles related to alcohol and drug related cases, he was counsel in Ex Parte Barry Wayne Buckner which redefined "driving under the influence of alcohol" in Alabama and in Lockard v. Town of Killen, the leading case defining a citizen's right to an independent blood test. He was also counsel in the case Mayo v. City of Madison, where in a landmark decision of the Alabama Supreme Court, the Alabama breath test program was thrown out. More information can be found at www.pricelawfirm.com



SPURGEON COLE, PH.D., is Professor Emeritus at Clemson University. He has published more than forty articles on tests and test evaluation. Dr. Cole has written one textbook and testified in numerous courts on issues relating to the admissibility of field sobriety test evidence. © 2001 by Phillip B. Price, Sr. and Spurgeon Cole, all rights reserved.

officers falsely arrested 24 percent of the innocent people using the SFST as the basis for the arrests.

San Diego, CA: 'Validation of the Standardized Field Sobriety Test Battery at BACs Below 0.10 Percent'¹¹

Anacapa Sciences, Inc. conducted the San Diego SFST "validation study" and reported the findings in August 1998. NHTSA funded the study, but it was authored by Jack W. Stuster, Ph.D., CPF, and Marcelline Burns, Ph.D. As the name implies, it was conducted in San Diego, California during 1997. They titled the study, "Validation of the Standardized Field Sobriety Test Battery at BACs Below 0.10 Percent." Seven officers of San Diego's alcohol enforcement unit tested 298 citizens and data was collected using BAC levels of .08 and .04 percent. *All officers had a PBT at roadside and no observers were used in the study.* They claim to have an accuracy in arrest/no arrest decisions of 91 percent at .08 BAC level and 80 percent using BAC levels of between .04 and .08 percent. They could have just as well have called this study a validation of the smell of alcohol, or blood shot eyes, etc. The researchers neither randomly selected subjects nor made any effort to isolate variance contributed by the SFST.

If you wanted to determine the effectiveness of the SFST, you would control all variables except the SFST. In this study maybe it was the smell of alcohol that accounts for the 7 percent increase over accuracy. We realize that the officer has the other things but they are claiming that the increase is due solely to the SFST. I had an uncle who drank too much. The next day he always was sick. He claimed that it was something he ate. Since he always ate something we will never know what made him sick. If we are to find out what made him sick, we would have him either drink or eat but not both. We will never know if the small increase in decision accuracy was due to the SFST or to the many other uncontrolled variables.

The fact that a PBT was furnished to the arresting officers with no observers present is an improper method of data collection. The manner in which data are collected during research studies is an integral part of the scientific process. Data must be collected in a trustworthy manner with objectivity built in to insure a fair sampling process. This so-called

"garbage-in/garbage-out" principal is fundamental to science. If one interjects subjectivity and/or the opportunity of unreliable data with no "controls," then the experiment fails. No reliable conclusions can be drawn from a study when all the participants are given a method and an opportunity to know the answers to "the test." There is a reason why double blind techniques are employed in research.

Psychological factors can and do influence behavior. If a doctor studying the clinical effect of a drug knew which subjects were taking the medication and which were taking the sugar pill, his or her ratings of improvement might be influenced. The doctor is not even aware that he is being influenced. In a recent study, schoolteachers were told that ten randomly chosen students had scored very high on intelligence tests. Even though the students were randomly chosen, their grades improved significantly over the next nine weeks.¹² The students IQs had not improved. The teachers were influenced by bogus information. The teachers were not cheating but were simply influenced by psychological factors. The seven officers in the San Diego study knew that the results of the study, and their effectiveness as trained DUI officers, would be known by practically every police department in the country!

If we provide the opportunity and the methodology of looking at the answers to the test to half of the class, we would expect that half of the class to score better than the other half which were not provided either the answers nor the opportunity to "cheat." If we couple this opportunity with the fact that the students who were given all of the answers to the test would have no supervision whatsoever, you would expect their half of class to fare even better. If you add even another variable in with our "answers-known" half, it becomes even more ridiculous. We're going to give the answers to half the class, not have *any* monitoring of the testing process, and let them *see* the correct answers *before* they fill out their tests! That is exactly what they did in the San Diego study. They gave the police officers PBT, provided no observers, and then had them fill out their data form that provided their "prediction" of the individual's BAC. This improper data collection procedure totally nullifies the attempted validation of the SFST.

With this type of data collection, it is

not surprising that the authors provided a "validity coefficient" for this study and not for the two studies that used observers. A proper "validation study" should indeed report the data with a validity correlation coefficient. It should be noted, however, that the San Diego study did not report a *reliability* correlation coefficient¹³ which is also required for a proper "validation." The validity correlation coefficient in the San Diego study was reported correlation coefficient of .65. The validity coefficient indicates the officers accuracy in predicting the subject's BAC level. A validity coefficient of .65 indicates that the error of estimate is 76 percent as large as it would be by chance. In other words, when officers had access to the BAC levels, their predictions were 24 percent better than by simply guessing the BAC level. In spite of this improper data collection, 29 percent of the sober subjects (below .08 percent BAC) were falsely arrested using the SFST.

Florida: 'A Validation Study Of the Standardized Field Sobriety (SFST) Battery'¹⁴

The Florida SFST "validation study" was conducted by the Institute of Police Technology and Management (IPTM), the Pinellas County Sheriff's Office, and the Southern California Research Institute (Marcelline Burns, Ph. D.). It was accomplished in cooperation with NHTSA. The researchers used the SFST. The study was conducted in 1997, however, the date of publication is unclear. Approximately two-thirds of the time, observers were present during the collection of the data. The study employed 313 subjects. Allegedly, the observers, and not the officers, had a PBT. The Florida study claims an overall accuracy rate of 88 percent.

The researchers provided neither reliability nor validity scores. Reliability and validity scores are necessary in order for the results to be properly evaluated by the scientific community. Legitimate tests such as the SAT, College Board tests, and The Wechsler Adult Intelligence Test, all report validity and reliability scores. The reliability scores for the above listed tests range from the low to middle 90s. The Florida study has the same base rate or "guess rate" problem as the Colorado study. The *mean* BAC of arrested drivers was .15 percent which, as the authors indicated, is a "severely impairing level."¹⁵

Eighteen percent had a BAC of between .200 percent and .284 percent. In this study, the officers had to have "some other evidence of impairment" before they even administered the SFST. Scientists use random sampling to control extraneous variables. None of three studies made any attempt to randomly select subjects. This oversight makes it impossible to generalize the results to other settings. The results of the three studies were seriously compromised or possibly robbed of all value by the confounding of variables. The researchers made no attempt to control such variables as the smell of alcohol, speech, or other cues of BAC in addition to the subject's performance on the SFST. Who knows? Giving the SFST may have actually *decreased* decision accuracy.

How do you think giving the SFST to a person weaving down the road, smelling like a brewery, with blood shot eyes, staggering out of the car, slurring his speech "improves" the guess that his BAC exceeds the .08 limit? They were looking for alcohol impaired drivers and only stopped drivers that they already believed to be intoxicated. In fact, approximately 80 percent of the detained drivers had a BAC above the legal limit. An accuracy rate of 90 percent does not look very good when you consider the guess rate is 80 percent and the mean BAC level is almost twice the "legal limit" of .08 percent. In addition, the officers already possessed significantly more information than simply the SFST results. A blind person could do this well.

Thirty-six percent of the time the officers had no observers present. Eighty percent of the subjects were men and twenty percent were women. Officers armed with the SFST arrested 18 percent of the innocent people involved in the study.

Discussion

All three studies seek to authenticate the notion that if the SFST battery is used by trained law enforcement officers, then the officers' arrest/no arrest decisions for DUI will be 88 percent accurate. NHTSA, by these studies, wants its readers to believe that if the SFST is given at roadside, the officer's decision to arrest or release an individual for DUI (DWI, OUI, etc.) will be correct 88 percent of the time. Moreover, NHTSA claims that this 88 percent correct arrest/no arrest decision is based *solely* on the SFST



Criminal Jury Instructions: Promoting the Defense Perspective

FORECITE® National

The Nation's most comprehensive
resource for criminal
jury instruction on a CD-ROM.

ORDER TODAY AND SAVE \$50.

Toll free 1-866-885-2483 or online
at www.forecite.com



results. The truth is they start with a base rate of 80 percent and make no attempt to isolate the effectiveness of the SFST. The 1977¹⁶ and 1981¹⁷ studies funded by NHTSA were conducted trying to employ sound scientific principles but the researchers did not obtain the desired results.

The only reason for the current pseudo-scientific studies is to try to convince the legal community that the new and improved SFST are valid. NHTSA wants this validation very badly. Unfortunately, these attempts at validation border on chicanery and have no legitimate scientific value. The choice of NHTSA's use of the word "validation" is not just happenstance. As we shall see below, in order for a scientific test to even be considered a "test" in the true sense of the word, the "test" must meet certain criteria. "Validity" is defined in terms of whether a test measures and/or predicts what it purports to measure or predict. For example, the pre-college entrance exam known as the SAT (Standard Achievement Test) is supposed to predict college student's grades. If scores on the SAT correlate with a particular student's grades, the SAT would have a degree of

validity proven over time. Validity is reported in terms of correlation coefficients. Coefficients can range from -1 to +1. Over the years, the SAT has proven to have validity in measuring what it purports to measure, *i.e.*, college grades. In the roadside DUI investigation, NHTSA is keenly desirous of obtaining some reliable and valid evidence that the SFST are predictive of blood alcohol levels. They long ago gave up hope that the SFST could predict driving impairment. Although they desire evidence of reliability and validity, two of the recent studies do not report *validity* coefficients and none report *reliability* coefficients.

One should realize at the outset that the researchers involved with these three studies were on a mission. They were hell bent on establishing validity regardless of what they had to do to find it. If using sound scientific technique could not find validity, then adjust the study. Do anything, but get the validity. NHTSA wants very badly to give these roadside coordination experiments some sort of technical blessing hoping to gain acceptance within the scientific community. If one were to accept the glossing overview given the studies by their authors, one might con-

clude the SFST have substantial validity, however, when the studies are examined closely, it is apparent that the Colorado, San Diego and Florida studies are so flawed that they are indeed useless. The truth is an officer has only a slightly better chance of making a correct arrest decision using the SFST than flipping a coin.

We will now discuss the three NHTSA sanctioned "validation studies" and determine first, whether the studies were done in a way consistent with accepted scientific experimentation. Second, whether the conclusions reached by the authors are valid interpretations of the data, and third, we shall reveal some of the "hidden" statistical data that neither NHTSA nor the authors discuss in the studies.

Experimentation and testing have been around about as long as civilization itself. It is an integral part of our existence here on earth. Scientific testing has evolved over the years and has acquired certain criteria that must be met before a particular test or experiment itself could be considered valid or worthwhile. The scientific community has adopted a set criteria by which a study is judged. It is often said, "the theory guides, but the experiment decides!" In order for the experiment, test or study to "decide," the experiment test or study must have been performed in a way to reasonably insure accurate and reliable results. What we are trying to do when we do an experiment, study or test is to first come up with a hypothesis. The hypothesis is a theory yet to be proven by sufficient fact(s). Galileo hypothesized that balls of various sizes would fall at identical speeds. Galileo's hypothesis was later supported by experimental data (he dropped balls from the leaning tower of Pisa). NHTSA hypothesized BAC levels could be predicted from scores on the SFST. The burden on NHTSA is to produce scientific data that support the hypothesis. Without supporting data, the SFST are reduced to laymen observations. Not being able to stand on one leg means only that you cannot stand on one leg. That is all that it means. If told that a person cannot stand on one leg, you can deduce nothing else from this piece of information.

NHTSA hypothesizes that if you give individuals the SFST, you will correctly arrest or not arrest for DUI 88 percent of the time. Given this hypothesis, then, the question is posed: Do these validation

studies support the hypothesis? In order to answer that question, we must look at the requirements of a valid test or study.

The purpose of a test, study or experiment is to predict a whole by extrapolating from a smaller sample. Anne Anastasi is a pre-eminent authority on the subject of test and measurement. Anastasi defined "test" as an objective and standardized measure of a sample of behavior.¹⁸ For example, if a researcher chooses to determine a person's word knowledge or math ability, he or she examines the individual's performance on a representative set of words or math items. The value of the tests depends on the degree that it is an indicator of a broader range of behavior. The original field sobriety tests measured samples of behavior and attempted to predict overall psychomotor skills necessary to operate an automobile. Researchers, however, found little or no relationship between field sobriety tests scores and driving skills. Dr. Burns and her cohorts decided that the new SFST could, however, predict the BAC level. She readily admits that the SFST do not predict or have anything to do with driving skills. The value of the scores on the SFST has no value unless it relates to BAC levels. Burns and NHTSA suggest that the SFST correctly predict impairment 88 percent of the time. If the studies were conducted properly, they might provide some evidence of validity, however, if they do not possess the necessary characteristics of a valid study, then no validity can be attributed to the studies. Let us look at the necessary criteria for a valid test. The three basic elements of a test:

1. Standardization
2. Reliability
3. Validity

Standardization. For an instrument to be considered a test, it must first be standardized. Standardization implies uniformity of procedures in administering and scoring the test. For scores to have meaning, tests conditions must be the same for all participants. The SFST have little or no standardization. They are frequently given in a variety of situations under varying conditions and with different instructions. The criteria for passing and failing often differs from officer to officer. For proper standardization, even voice inflection must be standardized. In addition, the scoring is often subjective

both between officers and within the same officer at different times. Supposedly, 20-year olds and 60-year olds are evaluated the same when it is obvious that the psychomotor skills of a twenty-year old differ tremendously from those of a sixty-year old whether the subjects are drinking or not drinking.

The SFST used to be only FST. In 1977 and 1981, NHTSA conducted lab studies in order to standardize the one leg stand, heel to toe and the nystagmus test. These three tests comprise the new and improved SFST.

When we speak of "standardization" within the context of this article, we must consider the term in two separate tiers or perspectives. The first use of the word "standardization" is within the context of whether each individual field sobriety test (FST) is given in a "standardized" way, *i.e.*, in the same manner. The other use of the term "standardization" in this article deals with whether the validation studies themselves were done in a standardized way in order to pass a threshold requirement to be considered one of the elements of a valid test or study. In the first context, *i.e.*, whether each individual FST was done in the standardized way, is not the subject of this article and is only mentioned here in passing. As we mentioned earlier the purpose of the 1981 NHTSA study was to "standardize" the three FST. NHTSA's use of the term "standardized" in the 1981 study is not by accident either. Standardization is the first step in the process of establishing a valid test or study. Standardization, or consistency, is an incipient phase of the scientific process. In the context of the individual test standardization, NHTSA does require that each of the tests be given in a precise manner for each test to have *any* validity.¹⁹ It is unclear whether or not all officers in these three studies used written "instruction cards" to insure all tests were instructed in exactly the same way. For purposes of this article, we will make the concession that differences may have existed, but is not a subject to be addressed here.

In our second context, the "standardization" issue becomes whether or not the validation studies themselves were carried out in a uniform, controlled way thereby not allowing variables to enter the equation. Non-uniformity in the testing process is the enemy of standardization, and non-standardized equals non-reliability. One of the arguments made in

the 2000 Presidential election in Florida was that in manually recounting election ballots throughout the state there was no statewide standard in determining the intent of the voter. The inclusion or not of so-called "dimple ballots" and "pregnant ballots" in the ballot counting process was a major issue in the litigation. There was a major concern in trying to establish a statewide *standardization* in those recounts. This lack of standardized method of counting ballots proved to be a major concern of the United States Supreme Court in its decision.

... it is obvious that the recount cannot be conducted in compliance with the requirements of equal protection and due process without substantial additional work. It would require not only the adoption...of *adequate statewide standards* (emphasis added) for determining what is a legal vote, and practicable procedures to implement them, but also orderly judicial review of any disputed matters that might arise.

Bush v. Gore, [No. 00-949, December 12, 2000] 531 U.S. _____ (2000).

Each of the three NHTSA validation studies were performed in three different ways with different criteria. So in terms of the three studies, they clearly were not done in a standardized way. Not only that, but within each of the studies themselves, extreme and important variables were present, thereby invalidating standardization of the studies. Of primary importance was the use of a PBT. The standardization requirement was not met.

As can be seen in the chart herein, none of the three studies were done in the same manner as the others.

	Colorado	San Diego	Florida
PBT use	None (except for observers)	100%	None (except for observers)
Observers Present	41%	None	64%
BAC Level	.05	.04 & .08	.08
Prior DUI Suspicion	80%		100%
Mean BAC	.152	.122	.15
Screened for Drugs	None	None	None

Differences also existed within each individual study, as well, thereby further invalidating this standardization requirement. Observers were only present part of the time, for example. No controls were placed on extraneous variables

entering into the arrest decisions. How many of the individuals arrested had slurred speech, how many staggered, how many admitted they were drunk, for example. With no controls on these variables (just to name a few), then each test itself cannot be said to be standardized. Even if one were to make the stipulation that the three tests were given in a uniform way in a particular study, there were still too many variables introduced in the studies to establish the standardization reliability requirement.

Reliability. The second requirement of a valid test is that the test must be proven to be reliable. If one gives a test in a standardized way, and one gets the same result each time, then the second prong of a valid test or study is provided, *i.e.*, reliability. Reliability basically means consistency or repeatability. A test or study, to be reliable must be given several times (in the same manner) and reach the same result each time. Without repeatability in the same manner, a test cannot be shown to be reliable. If variables are introduced in one study that were not used in the other one (or two), then the studies were not done in a standardized way and have neither the standardization nor the reliability prong established.

Let's say we are going to do a study to try to determine which tastes better, field corn or silver-queen corn. In order for the study to be reliable, the variables that the corn is subjected to must be the same (standardized). The test or study must be done more than once and reach the same result (reliable). To achieve this, we must plant both types of corn in the same type soil, with the same amount of sunlight, the same amount of water, etc. for there to be a fair comparison. If we provide the silver-queen corn with an irrigation system, it

would have, from the outset, an unfair advantage over the field corn. In the same vein, if we give the officers in the field (as NHTSA did in the San Diego Study) unfettered use of a PBT along with the SFST, then we have totally invalidated our attempt to validate. As noted in the 1983 NHTSA study, this use of PBT is unacceptable. As the Colorado study states,

BACs obtained with PBTs could have inflated their (arresting officers) accuracy rate, and the contribution of the SFSTs to their decisions (to arrest) cannot be unequivocally determined.²⁰

To quote the Florida Study:

During the study, the deputies did not have access to PBTs. That condition was necessary to insure that their decisions were based solely on general observations and the SFSTs.²¹

You will see that in the San Diego study officers in the field were supplied with a PBT. This clearly introduces a huge variable which, quite frankly, invalidates the validation study. Any data analysis from the San Diego study is meaningless. Giving a PBT to an officer in the field when the sole criteria is supposed to be the SFST, is like giving students the answer key and then sending the students to a private room to take the test with no teacher present. As noted above, the researchers noted in both the Colorado and Florida studies, as well as the NHTSA 1983 study, that allowing officers access to the BAC is unacceptable. Another contaminating variable was the fact that the subjects were stopped because the officer suspected they were driving under the influence of alcohol.

Validity. In order to have validity, reliability must be firmly established. Since there is neither standardization nor reliability, validity cannot be attained.

Base rate error. There is another type of error committed during these studies, and it is called "base rate error." The "base rate" refers to the frequency that a condition occurs in the sample room. Suppose one has the fear that his or her teenager is suicidal. Like all concerned parents, they take their teenager to a psychiatrist and ask the professional the likelihood of their teenager committing suicide? The parents are told that the teenager is tested and not to worry because the test is 99 percent accurate. Surely, a test that is 99 percent accurate is valid.

The truth is the test may have no validity. The base rate or as previously indicated as the "guess rate" is also 99 percent accurate. For example, if the psychiatrist simply predicts that no one will commit

suicide, he will be 99 percent accurate. The same base error occurred in the current so-called validation studies. For example, in the Colorado study, the base rate was 79 percent. Seventy-nine percent of the subjects had BACs above .05.²² In the Florida study, 81 percent of the subjects were above the target value of .08 percent. That means that if you arrested everyone you come in contact with in the Florida study, then you would make a correct arrest decision (of being over .08 percent) 81 percent of the time! If you arrested everyone with the smell of alcohol on his/her breath, you would probably be correct more than 80 percent of the time. In the Florida study, the subjects had to have some other evidence of impairment before they were asked to do SFST, thereby adding another variable and skewing the base rate even further. The decision to arrest was not based solely on the SFST.

Not only is there a "base rate error" present in all three studies, there are also some other generalized problems with the studies. The second major limitation is that the authors reported no validity or reliability scores for Colorado and Florida studies. Without validity or reliability scores, it is not possible to evaluate the effectiveness of the SFST. The authors report the results in the most elementary manner. A first-year graduate would be expected to provide a more comprehensive and sophisticated analysis.

The San Diego study contains correlation coefficients, but since PBTs were supplied to the officers with no observers present, the data are worthless. The San Diego study is an example of how data should be reported. Since neither the Colorado and Florida studies report either a validity or reliability scores, the results cannot be properly evaluated.

The reliability correlation coefficient is reported as a value from 0 to 1.0. A reliability coefficient of .70, for example, would mean that 70 percent of the score is "true score" and that 30 percent of the score is error score (false). For "test" purposes the reliability coefficient should be at least .90.²³ In the 1977 and 1981 studies, reliability scores were in the 50s and 60s. That is perhaps why they were not reported for the current studies. Since the same tests comprised the SFST in the current studies, there is no reason to believe that a miracle has occurred in the past 20 years. Without reliability and validity scores, it is not possible to calcu-

late standard error of measurements or standard error of estimates.

There has been no attempt to establish norms for the SFST. We have no idea how well a sober person can perform on the SFST. How does age or gender affect performance? How does fatigue or practice affect performance? If an individual performs poorly at a .11 percent BAC, how does that compare with his or her performance with a BAC of .00? Before any individual's performance can be considered a "test," that particular individual's baseline with no alcohol must be known and factored in. Without answers to these basic questions, the SFST remain in the same category as Tarot Cards.

The most important statistic derived from the studies from a defense perspective is the number of false arrests. This statistic is one that you have to "hunt" for in the study to find it. The authors, quite obviously, do not discuss this statistic. Of the sober individuals that were involved in the Colorado, Florida and San Diego studies, the officers falsely arrested 24 percent, 18 percent and 29 percent, respectively. That is an average of 23.6 percent false arrest rate. What this means is that if the SFST are used as a decision of whether to arrest individuals for an alcohol related offense, one out of every four sober people will be falsely arrested. While it might be considered a noble goal to rid our highways of intoxicated drivers, it seems to be a matter of diminishing return when one in four of us sober drivers (23.6 percent) will end up in jail using these SFST. This is totally unacceptable to think that NHTSA claims a 90 percent accuracy rate, but the innocent drivers are punished at a rate of 23.6 percent! NHTSA's claim of 90 percent is not a correct analysis of the data reported.

While, in an attempt to try to justify the false arrest rate in the Colorado study, the authors suggested that perhaps those individuals that were falsely arrested were impaired by other substances. The authors, by suggesting this, are totally invalidating all three studies because no one, in any of the three studies, was screened for drugs or any other impairing substance.

The Devil . . .

Neither one study nor any combination of the three validation studies supports the authors' conclusions. They want us to believe that officers employing the three tests battery is 90 percent accurate in

making decisions to arrest individuals at levels below .10 percent. The Devil is in the details. ■

DWI is devoted to presenting practical information on the defense of DUI/DWI cases. Authors and ideas are welcome. Please contact:

Steven Oberman
Daniel & Oberman
550 West Main Ave, Ste. 950
Knoxville TN 37902
Phone (865) 546-4292
Fax (865) 5464294
oberman@daolaw.com

NOTES

1. Anderson, Ellen and Marcelline Burns, Ph.D., *A Colorado Validation Study of the Standardized Field Sobriety Test (SFST) Battery*, November 1995.
2. *Colorado Validation Study*, 9.
3. *Colorado Validation Study*, v.
4. Sworn testimony of Marcelline Burns taken December 9, 1994 in *State of Florida v. William Meador, et al.*, Case Number 93-810MM10A (12-13), reported at 697 So.2d 826.
5. *Colorado Validation Study*, 14.
6. *Colorado Validation Study*, 16.
7. *Colorado Validation Study*, 17.
8. *Colorado Validation Study*, 21.
9. *Colorado Validation Study*, 19.
10. Price, Phillip B., Sr., *Fear and Field Sobriety*, 2000. See *THE CHAMPION*, Aug. 1996, p. 46.
11. Burns, Marcelline, Ph.D. and Jack Stuster, *Validation of the Standardized Field Sobriety Test Battery at BACs Below 0.10 Percent*, August 1998.
12. Rosenthal, R., *Experimental Effects in Behavioral Research*, New York: Appleton-Century Crofts, 1966.
13. *California Validation Study*, 26.
14. Dioquino, Sgt. Teresa, et al., *A Florida Validation Study of the Standardized Field Sobriety (S.F.S.T.) Battery*, (date of publication is unknown).
15. *Florida Validation Study*, 12.
16. Burns, M. and Moskowitz, H., *Psychophysical Tests for DWI Arrest*, DOT-HS-802 424, NHTSA, U.S. Department of Transportation, June 1977.
17. Tharp, V., Burns, M. And Moskowitz, H. *Development and Field Test of Psychophysical Tests for DWI Arrest*, DOT-HS-805 864, U.S. Department of transportation, NHTSA, Washington, 1981.
18. Anastasi, A., *Psychological Testing*, New York: Macmillan, 1998.
19. *DWI Detection and Standardized Field Sobriety Testing*, Student Manual, DOT-HS-178 R2/00, NHTSA, U.S. Department of Transportation., VIII-3.
20. *Colorado Validation Study*, 17.
21. *Florida Validation Study*, 8.
22. *Colorado Validation Study*, 14.
23. Anastasi, A., *Psychological Testing*, New York: Macmillan, 1998.

RIGHTS OF A PERSON REQUESTED TO SUBMIT TO A CHEMICAL ANALYSIS
TO DETERMINE ALCOHOL CONCENTRATION UNDER G.S. 20-16.2(A)
SPANISH VERSION (REVISED 12/97 BY P.O. JOE BLANCO)

Usted esta acusado de manejar un vehiculo en las carreteras o lugar de transito publico mientras cometiendo una ofensa de consentimiento implicado.

(EN OTRAS PALABRAS, MANEJAR UN VEHICULO BORRACHO.)

Le vamos a pedir que se someta a un analisis quimico para determinar la concentracion de alcohol en su cuerpo. Primero es requerido que usted este avisado oralmente y por forma escrita de sus derechos. Sus derechos son los siguientes:

1. Usted tiene el derecho de rechazar ser examinado.
2. Rechazar cualquier examen requerido resultara en una revocacion inmediata de su licencia de manejar un vehiculo por 30 dias por lo menos, y otra revocacion por 12 meses, hecha por la Division de Vehiculos de Motor.
3. Los resultados del examen quimico, o si usted rechaza a someterse al examen sera admitido como evidencia en las cortes por la ofensa que se le ha acusado.
4. Su licencia de manejar un vehiculo sera revocada inmediatamente por 30 dias si;
 - A. El examen indica una concentracion de alcohol de 0.08 o mas, o;
 - B. Si usted estuvo manejando un vehiculo comercial, y el examen indica una concentracion de alcohol de 0.04 o mas.
5. Usted puede tener una persona calificada que le administre el examen quimico, ademas de cualquier otro examen administrado por la direccion del policia.
6. Usted tiene el derecho de llamar a un abogado y de escoger un testigo que presencie los examenes, pero los examenes no se pueden demorar mas de 30 minutos despues del momento que se le aviso sus derechos.

DATE/TIME RIGHTS ADVISED _____

DID DEFENDANT CALL AN ATTORNEY/WITNESS? _____ TIME _____

WITNESS _____ OFFICER/DEPT. _____

CASE # _____ CITATION # _____

DEFENDANT'S NAME _____

DEFENDANT'S SIGNATURE(FIRMA) _____

SAMPLE OBTAINED BREATH _____ BLOOD _____ URINE _____