

THE SOBERING HISTORY OF FIELD SOBRIETY TESTS

Written by George L. Bianchi of The Bianchi Law Firm

605 Thomas

Seattle, WA 98109

Phone: (206) 728-9300

Fax: (206) 728-9305

Copyright: The Bianchi Law Firm 2004

I. THE AURA OF SCIENCE: GOVERNMENT STUDIES OF FIELD SOBRIETY TESTING

The results of field “sobriety tests,” long the mainstay of the case against a citizen accused of DUI, have for years been justifiably criticized as wholly subjective in their interpretation. The need to present such evidence in court as “scientific” was apparent, and the government responded. In June of 1977, the Southern California Research Institute (SCRI) was commissioned by NHTSA to study and evaluate the then currently used “field sobriety tests.” SCRI was given three mandates: (1) determine the alcohol sensitivity, if any, of such tests; (2) develop more sensitive and reliable tests, and (3) attempt to standardize the administration of “field sobriety tests.” The researchers were examining the tests to determine whether there was any link between the test results and intoxication or driving impairment. The goal was to develop alcohol-sensitive tests that would provide more reliable evidence by standardizing the tests themselves and the observations to be made. The result was Psycho-Physical Tests for DWI Arrests, DOTHS-802 424, (Burns & Moskowitz, June 1977).

Some of the original sixteen “tests” considered in the SCRI study included Alcohol Gaze Nystagmus, Walk-and-Turn, Romberg balance, Finger-to-Nose, One-Leg-Stand, and Finger Count or finger dexterity. Some of the tests used by police then and today, such as reciting the alphabet and counting backwards, were not ever considered.

In the quest to eliminate subjectivity from tests, the authors pursued the development of a “test battery.” The purpose of the battery was to provide statistically valid and reliable indications that a driver’s breath alcohol concentration (BrAC) level was not at or above 0.10. The battery was not designed to reveal indications of driver impairment.

Validation of the Standardized Field Sobriety Test Batter at BAC’s Below 0.10 Percent, DOTHS-___ at 28, (Stuster & Burns, August 1998). The result was a pilot program that studied a six-test battery (One-Leg-Stand, Walk-and-Turn, Finger-to-Nose, Finger Count, Horizontal Gaze Nystagmus (HGN), and Tracing) with three alternatives (Romberg balance, subtraction, counting backwards, and letter cancellation). The original data in the 1977 study suggested that it was unrealistic to attempt to use behavioral tests to discriminate subject alcohol level to a range of plus or minus 0.02 around the given BrAC level of 0.10 percent. Psycho-Physical Tests for DWI Arrests, at 41.

The task of rendering field sobriety testing “reliable” was daunting. The authors noted an error rate of 47% in arresting individuals who were under a BrAC of 0.10 percent. Id. at 28, 30, 102. Some of the sources of error were determined to be the failure of officers to heed the lack of test evidence, impairment which was not alcohol related, and officers who did not score the test properly. Id. at 28.

The study resulted in a three-test battery that included Horizontal Gaze Nystagmus, the One-Leg-Stand and the Walk-and-Turn test. The authors stated:

It became apparent during field visits that this objective [standardization of the tests and observation procedures] is highly important. There are wide differences between officers in using tests to assess a driver’s state of intoxication, and they may exist within the department as well as between agencies and locales. These differences seriously detract from reliability as well as from credibility of the officers in court proceedings.

Psycho-Physical Tests for DWI Arrests, at 59.

The three test battery was deemed “standardized” in 1981 with Development and Field Tests of Psycho-Physical Tests for DWI Arrests, DOTHS 805 864, (Tharp, Burns & Moskowitz, March 1981). The authors defined a standardized test as:

One which procedures, apparatus, and scoring have been fixed so that precisely the same testing procedures can be followed at different times and locations.

Id. at 3.

From August of 1978 until March of 1981 when the final report was concluded, Tharp, Burns, and Moskowitz worked to standardize the administration and scoring procedures associated with the three test battery (Walk-and-Turn test, the One-Leg-Stand test, and the Horizontal Gaze Nystagmus). Their results were evaluated in the laboratory and to a limited extent, in the field. Development and Field Tests for DWI Arrests, DOTHS 805, 864 (Tharp, Burns & Moskowitz, March 1981).

The authors found problems with the development of the field tests. For instance, for the Walk-and-Turn test, the authors noted that requesting people to “watch their feet” while performing this test increased its sensitivity to alcohol but at the same time made the test difficult for people with monocular vision (i.e., poor depth perception).

Performing the Walk-and-Turn tests with the eyes open and enough light to see some frame of reference was determined to be essential if sober individuals were to perform the test without difficulty. Id. at 4. It was found that some people have difficulty with the Walk-and-Turn test when sober, including people over 65 years of age, people with back, leg, or middle ear problems, and people with high-heeled shoes (over 2 inches). Id. at 5. The authors determined that the test required a line which the police officer could

manufacture. They also recommended that the Walk-and-Turn test be performed on a dry, hard, level, non-slippery surface and under relatively safe conditions. If those requirements could not be met at the roadside, it was recommended that the suspect be asked to perform the test elsewhere or that only the Nystagmus test be used. Id. at 5.

As it relates to the Gaze Nystagmus, the authors noted that approximately half of the “sober” people tested showed a slight Nystagmus in at least one eye when their eyes were deviated maximally. Id. at 7. The authors recommended that corrective lenses should be removed prior to the administration of this test. Id. at 7.

The researchers found that Gaze Nystagmus could be seen in 50% to 60% of all individuals if their eyes were deviated to the extremes and that Gaze Nystagmus occurs with some types of brain damage. Id. at 92.

It is difficult for an expert eye doctor, under the best of conditions, to detect and diagnose Nystagmus. But the “standardized” test battery calls upon a police officer to do just that in the field, under varying conditions. The researchers established precise directions for the Nystagmus test: to determine the onset of Nystagmus, the stimulus be moved fairly slowly (about 10 degrees per second) by the officer, otherwise normal oscillation of the eyeball may be mistaken for Nystagmus. Id. at 7. Further, on the second movement to of the stimulus in each direction, the recommendation was that the stimulus be moved faster (about 20 degrees per second) and that the observer should note (a) whether or not the suspect can follow smoothly and (b) how distinct the Nystagmus is at the maximum lateral deviation. Id. at 9. The authors found that the Gaze Nystagmus test may not be applicable to individuals wearing contact lenses, since hard contact prevent extreme lateral eye movements. Id. at 9.

What about alcohol and balance? The researchers noted that other variables such as sleep loss, increasing of room temperature, and eating can result in increased body sway, such as a police officer might attribute to alcohol. Id. at 83. Vision was also found to be important in relation to tests of balance and muscular coordination. The researchers opined, not surprisingly, that closing the eyes makes all of the balance tests much more difficult for both sober and intoxicated individuals. Id. at 83. The data, to them, suggested that the peripheral vision plays a particularly important role in maintaining balance. Id. at 84.

The end result of the 1981 study was an indication that the Gaze Nystagmus could correctly classify individuals at or above a BrAC of 0.10 seventy-seven percent (77%) of the time, that the Walk-and-Turn test could properly classify individuals as being at or above a BrAC of 0.10 sixty-eight percent (68%) of the time, and the One-Leg-Stand tests could properly classify individuals at or above a BrAC of 0.10 sixty-five percent (65%) of the time. When they combined the results of the Gaze Nystagmus with the Walk-and-Turn test, there was determined to be an eighty percent (80%) accurate classification of a person at or above a BrAC of 0.10. The authors also noted a thirty-two percent (32%) false arrest rate in the overall statistics.

In 1983, NHTSA commissioned Anderson, Schweitz, and Snyder to develop standardized practical and effective procedures for police officers to use in reaching an arrest/no arrest decision. Their research is found in Field Evaluation of Behavioral Tests Battery for DWI, DOTHS-806 475 (Anderson, Schweitz & Snyder, September 1983). The study tested the feasibility of using the three-test battery in operational conditions by police officers and would secure data to help determine if the three-test battery would

discriminate as well in the field as it had previously in the laboratory. Ultimately, this study mirrored the statistical results of the laboratory testing previously summarized in 1981 by Tharp, Burns, and Moskowitz in Development and Field Tests for DWI Arrest.

Jack Stuster and Marcelline Burns were commissioned by the National Highway Traffic Safety Institute to evaluate the accuracy of the standardized field sobriety test battery to assist officers in making decisions for DWI at alcohol concentrations below 0.10 percent. In August of 1998, their report was submitted to the National Highway Traffic Safety Administration. Validation of the Standardized Field Sobriety Test Battery at BAC's Below 0.10 Percent, DOTHS-____ (Stuster & Burns, August 1998). As the authors noted:

During the past sixteen years, NHTSA's SFSTs largely have replaced the invalidated performance tests of unknown merit that once were the patrol officers only in helping to make post-stop DWI arrest decisions. Regional and local preferences for other performance tests still exists, even though some of the tests have never been validated. Despite regional differences and 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 pre-arrest procedures for evaluating DWI in most law enforcement agencies.

Id. at 3.

Prosecutors who were interviewed for this study suggested that all law enforcement agencies should restrict their field-sobriety evaluations to the same standardized development of three tests. Id. at 24. The 1998 study showed that the Gaze Nystagmus test had the highest correlation of accuracy when compared to the actual measured breath test level of 0.08 percent. In this regard, the Gaze Nystagmus test

showed 65% correlation to the actual measured alcohol concentration level, with the One-Leg-Stand test showing a 45% correlation with the actual measured alcohol level. Id. at 17. Approximately 10% of the individuals were determined to be falsely arrested by law enforcement in that their alcohol level was estimated to have been greater than 0.08 percent, but later found to be below that level. Id. at 81. Worth discussing, however, is the range of BrAC that was not measured and correlated in the 297 individuals tested. The range of BrAC tested is from a 0.038 for eight underage females to a 0.07 for two underage females. There were no individuals whose measured BrAC level reflected 0.08 up to approximately 0.13 percent in the study. Thus, no individuals were tested at BrAC levels ranging from what appears to be a 0.08 up to an including a 0.13 which resulted in the above-listed statistical analysis. Id. at 16.

While recapping the history of the NHTSA studies on roadside testing, Stuster and Burns recognized the limitations on relevancy of the NHTSA standardized tests, stating:

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.

Validation of the Standardized Field Sobriety Test Battery at BAC's Below 0.10 Percent, DOTHS-___ at 28, (Stuster & Burns, August 1998).

As a result of the above studies, the National Highway Traffic Safety Administration published both student and instructor manuals to be used by law enforcement agencies for the detection and arrest of DWI suspects. DWI Detection and

Standardized Field Sobriety Testing. The first set of manuals were printed in 1981 with subsequent publications in 1992 (PB 94-780228 Student Manual, PB 94-780210 Instructor Manual), 1995 (AVA-19911BB00 Student Manual, AVA-19910BB00 Instructor Manual) 2000 (AVA-20839BB00 Student Manual, AVA-20838BB00 Instructor Manual) and most recently in 2002 (AVA-21135BB00 Student Manual, AVA-21134BB00 Instructor Manual). These manuals provided to law enforcement do not incorporate the statistical results of Stuster and Burns wherein they attempted to validate the three tests SFST batter to alcohol levels below 0.10 percent. The manuals incorporate and instruct law enforcement on the statistical results of the three studies leading up to Field Evaluation of Behavioral Tests Batter for DWI, DOTHS-806 745 (Anderson, Schweitz & Snyder, September 1983).

II. PROCEDURES: INSTRUCTIONS FROM OFFICIAL FST MANUALS

The National Highway Traffic Safety Administration has defined DWI detection

The entire process of identifying and gathering evidence to determine whether or not a suspect should be arrested for DWI violation.

SFST Student Manual (2002) (AVA-21135-BB00) at IV-1; SFST Instructor Manual (2002) (AVA-21334-BB00) at IV-1.

DWI detection is divided into three phases: (1) Vehicle in Motion; (2) Personal Contact, and (3) Pre-Arrest Screening. SFST Student Manual (2002) (AVA-21135-BB00) IV-2-5; SFST Instructor Manual (2002) (AVA-21334-BB00) at IV-1-7.

Phase One, Vehicle in Motion, involves observing the vehicle in motion and deciding whether there is sufficient cause to command the driver to stop. SFST Student Manual (2002) at IV-2-5, V-1-11, SFST Instructor Manual (2002) at IV-1-7, V-1-7.

Naturally, following this phase is Phase Two, Personal Contact. At this time, the officer is to observe and interview the driver, face-to-face, in order to decide whether there is sufficient cause to instruct the driver to step from the vehicle for further investigation. SFST Student Manual (2002) at IV-2-5, VI-1-6; SFST Instructor Manual (2002) at VI-1-11. It is at this point the officer makes observations to determine whether or not it is appropriate to order the individual to exit his/her vehicle to perform standardized field sobriety testing. Then comes Phase Three, Pre-Arrest Screening. Here to officer will determine if there is probable cause to arrest the suspect for DWI by the use of the standardized field sobriety testing (psycho-physical tests) which have been identified and “validated” through NHTSA’s research program. SFST Student Manual (2002) at VII & VIII; SFST Instructor Manual (2002) at VII & VIII.

Proper training of law enforcement officer under the approved curriculum consist of 16 sessions that span 22 hours, 45 minutes of instruction, excluding breaks. While NHTSA recognizes there may be some need of flexibility in the curriculum, they state that:

All of the training objectives are considered appropriate and essential for police officers who wish to become proficient at detecting evidence of DWI and at describing that evidence in written reports and verbal testimony. All of the subject matter is considered necessary to achieve those objectives. All of the learning activities are needed to ensure that the students master the subject matter.

This course is “flexible” in that it can easily be **expanded** since it does not cover all dimensions of DWI enforcement.

SFST Instructor Manual (2002) at 7.

**THE STANDARDIZED FIELD SOBRIETY TESTS
ARE NOT AT ALL FLEXIBLE. THEY MUST BE**

**ADMINISTERED EACH TIME, EXACTLY AS
OUTLINED IN THIS COURSE.**

SFST Instructor Manual (2002) at 8.

**IT IS NECESSARY TO EMPHASIZE THIS
VALIDATION APPLIES ONLY WHEN:**

- **THE TESTS ARE ADMINISTERED IN THE
PRESCRIBED, STANDARDIZED MANNER**
- **THE STANDARDIZED CLUES ARE USED TO
ASSESS THE SUSPECT'S PERFORMANCE**
- **THE STANDARDIZED CRITERIA ARE
EMPLOYED TO INTERPRET THAT
PERFORMANCE.**

**IF ANY ONE OF THE STANDARDIZED FIELD
SOBRIETY TEST ELEMENTS IS CHANGED, THE
VALIDITY IS COMPROMISED.**

SFST Student Manual (2002) at VIII-19.

Unfortunately, many police agencies approve deviations from these approved and validated field sobriety test procedures without challenge from the defense.

III. FIELD SOBRIETY TEST EVIDENCE IS INADMISSIBLE WITHOUT A PROPER FOUNDATION

In Washington State, it is not illegal to have consumed alcohol and driven a motor vehicle. It becomes a crime only when the alcohol affects someone's ability to drive to any appreciable degree. See, State v. Hurd, 5 Wn.2d 308, 105 P.2d 59 (1940) and State v. Hansen, 15 Wn.App 95, 546 P.2d 1242 (1976).

In the absence of foundation testimony establishing the reliability and relevance of field sobriety tests and physical observations to show alcohol-induced impairment of the ability to drive a motor vehicle, such test results and observations should be excluded

from use as evidence. In any trial which **does not involve a breath test**, the government sponsored studies discussed in this article stand for the proposition that the standardized field sobriety tests **are not relevant** and should be inadmissible in the DUI trial.

Additionally, FST evidence must be shown to be relevant (ER 401) and more probative than prejudicial (ER 403) before it may be considered by the trier of fact. Probative value is not shown without evidence of a physiological relationship between the consumption of alcohol and the decreased ability to perform the specific physical tests requested by the officer. No study supports this assertion. In the absence of such showing, the prosecution should be precluded from making any reference to these tests. In the absence of a proper foundation, the FSTs are either irrelevant or unduly prejudicial.

ER 104(b) requires:

When the relevancy of evidence depends upon the fulfillment of a condition of fact, the court shall admit it upon, or subject to, the introduction of evidence sufficient to support a finding of the fulfillment of the condition.

The comment section (b) of ER 104 is helpful to the analysis:

This section is the same as ER 104(b) and defines a procedure for handling the situation in which a party wishes to prove Fact A, but Fact A is relevant only if Fact B is established. The order of proof under this Rule, as generally, is determined by the Judge. Rule 611. The Court, in its discretion, may decide whether to hear evidence of Fact A or B first, taking into account the relative prejudice of having the jury hear one rather than the other if the proponent fails to offer evidence of one of them sufficient to warrant a finding of its truth. Because of this danger of prejudice, the Rule should be used with caution, especially in criminal cases.

Using the analysis of Evidence code comment, “Fact A” in a DUI case is that the defendant’s allegedly poor performance on the field sobriety tests is a result of and

therefore proof of his/her intoxication. “Fact B” would be that the field “sobriety” tests demonstrate a reliable, trustworthy, and recognized relationship between the consumption of alcohol, the ability to perform the specific field “sobriety” tests properly and driver impairment. Under the analysis of the ER 104, the prosecution is required to lay the proper foundation (Fact B) before any reference is made to the results of field “sobriety” tests (Fact A).

Any testimony that infers field sobriety testing demonstrates an individual’s ability to drive a motor vehicle is impaired or affected to an appreciable degree should not be allowed unless it can be said with substantial assurance that the presumed fact (driver impairment) is more likely than not to flow from the proven fact (field sobriety tests) on which it is made to depend. c.f. State v. Kovac, 50 Wn.App 117, 120, 747, P.2d 484 (1987) quoting from Turner v. U.S., 396 U.S. 398, 405, 90 S.Ct. 642 (1970).

In the past, trial courts and case law have allowed for the inappropriate perpetuation and perseveration of relevance and reliability associated with field sobriety tests without any foundation whatsoever. The recent case of State v. Baity, 140 Wn.2d 1, 991 P.2d 1151(2000) suggests that the courts in Washington recognize that field sobriety testing in the state of Washington has gone through significant changes and testing to determine their validity and reliability. The Supreme Court in Baity determined that if any evidence of testing by an officer as it relates to DRE is to be allowed (which includes One-Leg-Stand, Walk-and-Turn, and AGN), the individual must have gone through the appropriate training and conducted the tests in accordance with the NHTSA standards.

In the case of U.S. v. VanGriffen, 874 F2d 634 (9th Cir. 1989), it was determined that the manuals as it relates to field sobriety testing were relevant and could be used as a

party admission under ER 801(d)(2). Further, the court in State v. Meador, 674 So.2d 286 (Fla.App. 4 Dist. 1996), ruled that any reference to the “exercises” by using terms such as “test”, “pass”, “fail”, or “points” creates a potential for enhancing the significance of the observations in relationship to the ultimate determination of impairment, giving them an aura of scientific validity and therefore, such terms must be avoided to minimize the danger that a jury will attach greater significance to the results of the “field sobriety exercises.” Id. at 833. The court also noted that the defendants:

. . . do raise genuine concerns about the scientific validity and reliability of the field sobriety exercises in predicting impairment. The studies conducted by NHTSA revealed that there is no reliable numerical correlation between performance on the field sobriety tests and breath alcohol concentration, let alone impairment.

Id. at 832.

In the case of State v. Homan, 89 Ohio St.3d 421, 732 N.E.2d 952 (2000), the Ohio Supreme Court ruled that for the results of any field sobriety tests to serve as evidence of probable cause to arrest an individual for driving under the influence, the police must have administered the tests in compliance with the standardized test procedures by NHTSA. The Ohio Supreme Court determined:

The HGN test is not the only field sobriety test that requires special care in its administration.

Id. at 425. The court went on further to state that there must be strict (not just “substantial”) compliance by the law enforcement officer with the NHTSA standards. In stressing the need for strict compliance, the Ohio Supreme Court stated:

. . . In the substantial-compliance cases, the minor procedural deviations that were at issue in no way affected the ultimate results. In contrast, it is well established that in field sobriety testing even minor deviations from the standardized procedures can severely bias the results.

Moreover, our holdings in the substantial-compliance cases were grounded, at least in part, on the practical impossibility of strictly complying with the applicable administrative regulation. In contrast, we find that strict compliance with standardized field sobriety testing procedures is neither unrealistic nor humanly impossible in the great majority of vehicles stops in which the police choose to administer the tests.

Homan at 426.

Defense attorneys should always question the relevance of field test evidence and move for exclusion of the tests in light of the numerous scientific studies referenced in this article. The tests are not admissible without foundational evidence. Until the required foundational training and background of the arresting officer is presented under ER 702, a defendant's motion to exclude or suppress the results of standardized FST or any such tests, as well as any reference to them as "sobriety tests," must be granted. The NHTSA commissioned studies have proven one fact beyond dispute, however, the term field "sobriety" tests is a misnomer.