September 1983
NHTSA Technical Note

DOT HS-806-475

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

Field Evaluation of a Behavioral Test Battery for DWI

Research and Development

Office of Driver and Pedestrian Research
Problem-Behavior Research Division

FIELD EVALUATION OF A BEHAVIORAL TEST BATTERY FOR DWI

Theodore E. Anderson

Robert M. Schweitz

Monroe B. Snyder

This paper presents initial findings from a recently conducted field evaluation of a sobriety test battery. Police officers from four jurisdictions were trained in the use of the sobriety test battery. They then administered the battery to drivers stopped for suspicion of Driving While Intoxicated (DWI) during the three month test period. The results indicate that the test battery can be easily administered in the field and is effective in determining whether a driver's Breath Alcohol Concentration (BAC) is above or below .10%.

I. Background

Estimates suggest that alcohol is involved in a large proportion of the fatal and injury accidents nationwide. Current attempts to deter the drinking driver are directed at raising the perceived risk of arrest and punishment. Unfortunately, research indicates that there is a very low actual risk of arrest, and the public's perceived risk is also quite low.

One factor that may contribute to the low probability of a drinking driver being arrested for a DWI trip is the difficulty police officers have in discriminating those drivers with BACs above 0.10% who are not obviously impaired. As a rule, police officers seem reluctant to arrest a driver unless there is a high degree of certainty that the drinking driver's BAC is above 0.10%. This results in the arrest of only those drivers whose impairment is quite clear and unquestionable. It has been estimated[1] that there are three times as many drivers on the road with BACs in the 0.10% to 0.14% range as in the 0.15% to 0.19% range. However, at least twice as many drivers are arrested who have a BAC in the 0.15% to 0.19% range as there are drivers arrested with BACs in the 0.10% to 0.14% range.

Every State in the country has either a "presumptive" or "illegal per se" law that makes reference to a BAC level, typically 0.10%. As a result, police officers have found it difficult to get a conviction for a driver whose BAC is less than 0.10%, or sometimes even close to it (unless other behavioral evidence is strong). The low level of detection and arrest of drivers with BACs only slightly above 0.10% may be the result of the lack of effectiveness of the techniques used by the officer in the field, who must make the initial determination regarding the driver's impairment level.

During a typical DWI investigation, the police officer who has formed an initial suspicion that a driver is impaired by alcohol, will sometimes administer a series of behavioral tests to the driver. These tests serve to confirm the initial suspicion and may provide probable cause to arrest the driver for DWI. Also, the driver's performance on these behavioral tests is sometimes a critical part of the evidence presented in court to support the DWI charge. At present, the tests and procedures used vary between local agencies and officers. For many of these tests, the relationship between performance and specific BAC levels has not been well documented. Thus, the National Highway Traffic Safety Administration (NHTSA) undertook a program to develop a behavioral test battery that is empirically related to BAC level and that will assist police officers to discriminate BAC levels more effectively.

An initial study[2] reviewed various tests that were or could be used for this purpose. Six tests were evaluated in a laboratory study. Three were recommended for development as a test battery that could be administered by police officers at the roadside. A second study[3] standardized the procedures for administering and scoring each test and collected data on their effectiveness in a controlled setting. The three tests are:

1. One Leg Stand. This test requires that the subject stand on one leg for approximately 30 seconds. The time requirement is important, because it makes the test sensitive to drivers with BACs in the 0.10% to 0.15% range, who may pass the test if they only have to balance for 10 to 20 seconds.
2. Walk and Turn. This is given in two parts. The first part requires that the subject balance heel-to-toe while listening to the instructions. In other words, the subject must do two things at once - balance heel-to-toe and listen to the instructions. Doing two things at once is very difficult for an intoxicated person. The second part of the test requires that the subject take nine heel-to-toe steps long a line, turn around, and take nine heel-to-toe steps back.
3. Gaze Nystagmus. Nystagmus means a jerking of the eyes. Gaze nystagmus refers to a jerking of the eyes as they gaze to the side. Many people will exhibit some nystagmus, or jerking, as their eyes track to the extreme side. However, as people become more intoxicated, the onset of the nystagmus, or jerking, occurs after fewer degrees of lateral deviation, and the jerking at more extreme angles becomes more distinct.

The ability of the sobriety test battery to assist police officers in determining whether the BAC of a person stopped for suspicion of DWI was above or below 0.10% was tested under laboratory conditions. A total of 441 subjects were dosed to varying BAC levels (between 0% and 0.19%) and scored, by participating police officers, according to their performance on each of the three sobriety tests. Given the knowledge of the subjects performance and scores on each test, the police officers correctly classified 81% of the subjects as being at or below .10%. Nine percent of the subjects were classified as above .10% although they were actually below .10%. Ten percent were classified as below .10%, although they were actually above .10%. One should also remember that the percentage of correct classifications will depend on the BAC levels of the subjects. The lab study attempted to get a range of BACs but did not get representation of the distribution of BACs that an officer might encounter at the roadside.

Although the police officers in the second study did use standard procedures for administering each test, they did not use a standardized procedure for combining results and reaching an arrest/no arrest decision. Standard procedures for interpreting combined results should optimize the effectiveness of the battery and strengthen the use of the results in court.

II. Study Objectives

The objectives of the current study were to:

- develop standardized, practical and effective procedures for police officers to use in reaching an arrest/no arrest decision when giving one or more of the three sobriety tests;
- test the feasibility of use in operational conditions by police officers; and
- secure data to help determine if the tests will discriminate about as well in the field as in the lab.

III. Analysis and Development

Laboratory data from the Psychophysical Tests Development Study[3] were used to develop procedures for police use in drawing conclusions from test results. The objective was to have procedures that:

- were quick and easy to use;
- could be used whether the officer decided to give one, two or three of the tests; and
- would maximize the detection of drivers at BACs of .10% or above while minimizing the continued investigation of persons below .10% BAC.

Various scoring procedure were examined that combined the results of the three sobriety tests. The procedure that was best able to classify the laboratory subjects with respect to their BAC levels was one that combined the Gaze Nystagmus and Walk and Turn test scores. A table was developed for use with this procedure that contains Walk and Turn test scores as row entries and Gaze Nystagmus test score: as column entries (see Figure I). Some of the boxes in the table are darkened. If the box at the intersection of a subject's Gaze Nystagmus and Walk and Turn test scores is darkened, then the subject's BAC is predicted to be at least 0.10%.

FIGURE I
Combined Test Scoring Procedure

GAZE NYSTAGMUS TEST SCORE

Using this procedure with the laboratory data, and an estimate of the BAC distributions expected for persons stopped by police officers, the expected accuracy of correctly classifying subjects as above or below .10% was 80%.

Individual cutoff scores were identified for each test, if it was the only one used, so as to maximize correct classification above or below .10%. The scores and estimated accuracy for the population expected to be encountered in the field are as follows:

*   Gaze Nystagmus - (Expected Accuracy - 77 percent) - If the test score is greater than 3, classify the subject as having a BAC above 0.10%.

*   Walk and Turn - (Expected Accuracy - 68 percent) - If the test score is greater than 1, classify the subject as having a BAC above 0.10%.

*   One Leg Stand - (Expected Accuracy - 65 percent) - If the test score is greater that 1, classify the subject as having a BAC above 0.10%.

"an estimate of the BAC distributions expected..." is another way of saying, "the expected prevalence..."

But they don't say what prevalence they expected!! See: Bayes' Theorem

IV. Field Evaluation

Four police agencies participated in the three month field evaluation. They were Arlington County (Virginia) Police; Maryland State Police; North Carolina State Police; and Washington, D.C. Police. The test period lasted from November 15, 1982 thru February 15, 1983. Due to legal problems surrounding the use of the evidential breath test device in Virginia, the Arlington County Police were forced to limit their field data collection period to two months.

A. Training

Training sessions were conducted at each of the police agencies during early November, 1982. Each police officer participating in the field evaluation attended a one day training session and was given a training manual that included the newly developed scoring procedures. The manual also covers the history and purpose of the standardized field sobriety test battery and administrative procedures including conditions under which the tests must be administered to be considered valid.

The first part of the training session was devoted to reading and explanation of the training manual. Next, the participants viewed a videotape. It demonstrated how to administer and score the sobriety battery and then gave the trainees an opportunity to practice their newly acquired skills by showing several subjects being given the three tests. Lastly, the police officers received instruction in how to present the behavioral data when testifying in court.

The second part of the training session was devoted to practice. Several volunteers (not participants) were dosed to BAC levels between 0.08% and 0.16%. The trainees then practiced administering the sobriety tests to the dosed volunteers. Their performances during this phase of the training session were critiqued by the course instructor.

B. Data Collection

Police officers participating in the field evaluation were requested to administer the sobriety battery tests to all persons they stopped for suspicion of DWI during a three month period. This was done in conjunction with their normal DWI arrest. They were asked to administer and score the sobriety battery tests prior to using a preliminary breath testing (PBT) device. The reason for this ordering was to reduce the possibility that the police officers' scoring of the sobriety tests might be influenced by the BAC results obtained from the PBT device. They were also asked to record the following data for each DWI stop made:

* Date of suspected DWI stop
* Where the sobriety test battery was administered
* Gaze Nystagmus score
* Walk and Turn score
* One Leg Stand score
* Angle of onset of the nystagmus
* Officer's estimate of the suspect's BAC
* PBT result (except North Carolina where PBTs are not used)
* Arrest disposition
* Evidential BAC result (if the suspect was arrested for DWI)

If the evidential BAC results were not available at the end of the shift, then they were added to the data form as soon as they became available.

Efforts were made to secure data for all DWI traffic stops for all tests and to minimize the possibility that knowledge of PBT results would be available to officers before administering or recording battery scores. However, the data were collected in operational situations where the first priority was law enforcement and public protection rather than research data collection. It was not possible for researchers to routinely accompany the patrols and supervise or observe the actual data collection. Therefore, no statements can be made as to how closely the requested data collection procedures were followed.

On a few occasions, NHTSA researchers rode along with police officers during their normal duty tours and observed them administering and scoring the sobriety battery. The purpose of this procedure was to determine whether sobriety battery tests were being scored according to the standardized instructions and to assist the police officers in perfecting their testing techniques.

There were several other major sources of data collected during this project. All participating police officers were surveyed before the sobriety battery training session and after the completion of the three month usage period to determine their opinion of the utility of the sobriety battery.

The cooperating police departments agreed to collect DWI arrest data for a three month period prior to the field evaluation for use as comparison data. Also, court dispositions for the DWI arrests both before and during the field evaluation are to be collected as they become available.

In two of the police agencies (Washington, D.C. and North Carolina) control groups were established for comparison purposes. These officers were not trained in the use of the sobriety battery, but were requested to fill out information forms on each DWI stop made during the three month field evaluation period. The data they supplied were similar to that supplied by the specially trained police officers, with the exception of the sobriety battery test results.

V. Results

Since the DWI arrest data for the three month period before use of the test battery, and the court disposition data have not yet been received, only the data collected during the three month field evaluation period are presented.

Some of the analysis involving BAC information used the preliminary breath tester (PBT) data and some used the evidential breath tester (EBT) data. Although EBT data were more precise, they are available only for arrested drivers. When BAC data were needed for as many drivers as possible who were stopped for suspicion of DWI, PBT data were used. Since the North Carolina State Police do not use PBTs, analyses using PBT results are based only on data from the other participating police agencies.

During the field evaluation (November 15, 1982 thru February 15, 1983) battery-trained police officers recorded data on the following number of drivers that they stopped for suspicion of DWI:

*   Arlington County Police - 345 (Note: Arlington did not record data on suspected DWI stops made after early January, 1983)

*   Maryland State Police - 451

*   North Carolina State Police - 434

*   Washington, D.C. Police - 276

During this same period of time officers in the North Carolina State Police control group recorded data on 813 drivers stopped for suspicion of DWI, and those in Washington, D.C. recorded data on 195 drivers stopped for suspicion of DWI.

Table 1 shows the percent of drivers stopped for suspicion of DWI that were given each test as well as the percent that were given all three of the sobriety battery tests. (PBT usage is also shown in Table 1.)

TABLE 1
Sobriety Battery Test and PBT Usage
by Police Agency

Police Agency	Gaze Nystagmus	Walk & Turn	One Leg Stand	All Three Tests	PBT
Arlington County Police	84%	76%	72%	70%	92%
Maryland State Police	92%	91%	90%	88%	63%
North Carolina State Police	91%	85%	85%	82%	0
Washington, D.C. Police	82%	78%	76%	74%	87%
All Police Agencies	89%	84%	82%	80%	—
Washington, D.C. Police - Control	0	0	0	0	94%

The percent of drivers that were given all three sobriety tests varies from a low of 70 percent for the Arlington County Police to a high of 88 percent for the Maryland State Police. The average usage rates for all Police Agencies were 80 percent for the complete sobriety test battery, 89 percent for the Gaze Nystagmus, 84 percent for the Walk and Turn, and 82 percent for the One Leg Stand. PBT use exceeded the use of the behavioral tests except in Maryland.

Table 2 documents the resulting accuracy of the Combined Testing Procedure (Gaze Nystagmus and Walk and Turn tests) and the three individual sobriety battery tests. Accuracy refers to the test's ability to correctly classify the suspect's BAC as above or below .10% (using PBT data). As indicated in Table 1, the PBT was not given to all the drivers stopped by the police. Therefore, the accuracy figures in Table 2 cannot be considered as applying to the entire population of drivers expected to be stopped by the police on suspicion of DWI.

TABLE 2
Accuracy of the Behavioral Test Scoring
Procedures in Predicting BACs

Police Agency	Two Test Combination	Gaze Nystagmus	Walk & Turn	One Leg Stand
Arlington County Police	76%	75%	72%	72%
Maryland State Police	96%	96%	94%	92%
Washington, D.C. Police	75%	73%	73%	73%
All Police Agencies	83%	82%	80%	78%
Estimated from Lab Data	80%	77%	68%	65%

The accuracy of the Combined Procedure for all Police Agencies (83 percent) compares favorably with the 80 percent accuracy computed from the laboratory data. Of the misclassifications; 16 percent involved classification of a driver's BAC as greater than or equal to 0.10% when his/her BAC was less than 0.10%; and 1 percent involved classifying a driver's BAC as less than 0.10% when his/her BAC was greater than or equal to 0.10%. Also the ranking, with respect to accuracy, of the four scoring procedures remained the same as that obtained from the laboratory data, i.e., the relative ranking from most accurate to least accurate was Combined Procedure, Gaze Nystagmus, Walk and Turn, and One Leg Stand. However, the differences in accuracy among the three tests were less than in the previous laboratory study. There are two differences between the lab and field studies that may explain the somewhat different results (e.g., improved accuracy especially for walk and turn and one leg stand tests). First of all, the instructions regarding the interpretation of subjects performance scores were modified and were specific and definite about what scores indicated a DWI. The second difference is the BAC distribution of the subjects who were tested. We do not know the distribution for subjects stopped, nor for those tested, but only for those who were give a PBT or arrested and given an EBT. Therefore, it is difficult to estimate how important the difference in BAC distribution may be in accounting for the observed accuracy improvements.

The data in Table 2 should NOT be used to draw conclusions about the precise accuracy of using only one given test by itself as opposed to using another one of the three by itself. The main reason is that in most cases, all three

tests were given in the same order with gaze nystagmus first. The results of the gaze nystagmus test were then known to the officer and may have had some subtle influence on his expectations and scoring of the next two tests.

Two major reasons make it necessary to be extremely cautious in analyzing the data collected in this study to draw conclusions about the relative effectiveness of the different techniques that were used. First of all, officers were not randomly assigned to different groups and differences in outcomes may be due to selection and assignment bias. Second, the only effectiveness data available in this study relates to the BAC distributions for subjects who were arrested, and for some others who were given PBTs. There are a number of problems in using these data. We do not know how those given a PBT differ from or are representative of the rest. Perhaps most significant of all, except for North Carolina, all agencies had PBTs available, and in the great majority of the cases, PBT data were available to the officer for a driver before he was arrested. Thus, most arrest decisions were based on PBT data, rather than just test battery data. Given these limitations and constraints, a few additional analyses were done that can be used to help compare and assess the different DWI detection techniques.

Table 3 presents data on the BAC distribution for drivers arrested as a result of police use of different procedures. The BACs are based on EBT results. The percent of arrested subjects falling in each BAC range is presented in the body of the table, for each different procedure. The procedures are as follows: (1) PBT and Normal Police Procedures. This was the Washington, D.C. control group, that did not use the sobriety test battery, but did use PBTs (in 94% of the stops). (2) Sobriety Test Battery and PBT. This procedure was used by the D.C., Maryland and Arlington police who had been trained in the test battery. (3) Sobriety Test Battery, no PBT (NC); arrest indicated by 2 test combined decision rule only. These data are based on arrests made by the North Carolina State police who were trained in the use of the test battery. No PBTs were available. Only those cases for which the combined 2 test score indicated there should be an arrest were included in this data set. (4) Sobriety Test Battery, no PBT (NC); officer arrest decision. This was similar to (3) above but also included cases in which the officer decided to arrest even though the combined two test score indicated no arrest. (5) Normal Procedures, no PBT (NC). This was the North Carolina control group which had neither PBTs or the sobriety test battery available.

Table 3 presents BAC data (based on EBTs) in 3 categories of operational relevance to the police. BAC category 1 (0 - .04) contains obvious false positives (people who are not legally impaired due to alcohol, but are arrested). However, it should be noted that some or all of these people may have been impaired from drugs other than alcohol. The information required to assess the extent of this factor was not available. Category 2 (.05 - .10) contains people who may be impaired - legally as well as in their performance; however, the BAC by itself will not prove it. Whether people in this category were good arrests or poor ones cannot be determined with the data available. Category 3 (.10+) contains people who would be considered legally impaired, even in the absence of signs of behavioral impairment, in States with "per se" legislation.

Table 3 shows relatively little difference between the resulting BAC distributions for police using PBTs and the test battery or the test battery alone. However, use of the PBT and/or test battery appears far superior when compared to the normal DWI arrest procedure.

TABLE 3
Percent in Each BAC Category for Drivers
Arrested by Various Procedures*

Procedure	False Positive 0 - .04%	Difficult To Assess Depends on Other Data .05 - .09%	Arrest Supported By BAC Data .10%+	N
1. Normal Procedure Using PBT (D.C. Control)	0	10	89	(164)
2. Sobriety Test Battery and PBT (D.C., MD & Arlington)	2	8	90	(581)
3. Sobriety Test Battery, No PBT (NC); Arrest Indicated by 2 Test Combined Decision Rule Only	4	11	86	(279)
4. Sobriety Test Battery No PBT (NC) Officer Arrest Only	4	12	83	(289)
5. Normal Procedures, No PBT (NC)	26	15	59	(309)

* some rows do not add to 100 due to rounding

Table 4 presents information on the BAC distribution for arrested drivers where the arrest decision was indicated by two of the sobriety test scores (no PBT available). It shows that when both the Walk and Turn and Gaze Nystagmus recommended arrest, 92% of the subjects were above .10%. If the two test combination and the gaze nystagmus score by itself recommended arrest, even though the Walk and Turn recommended no arrest, 77% were above .10%. Finally, if the walk and turn by itself and the combined score recommended arrest even though the gaze nystagmus score by itself recommended no arrest, 53% were above .10%.

Table 4
Percent in Each BAC Category
for Arrested Drivers Given
Two Sobriety Tests

Arrest Recommended by:			Resulting BAC Distribution			
Walk & Turn	Gaze Nystagmus	Two Test Combination	0-.04%	.05-.09%	.10%+	N
Yes	Yes	Yes	4	4	92	(74)
No	Yes	Yes	15	8	77	(13)
Yes	No	Yes	23	23	53	(13)

VI. Conclusions

The results of the field evaluation:

Confirm the laboratory findings regarding the ability of the sobriety test battery to effectively discriminate between drivers with BACs less than 0.10% and drivers with BACs over 0.10%.

Demonstrate that the three sobriety battery tests (Gaze Nystagmus, Walk & Turn and One Leg Stand) can be easily and effectively used in the field by police officers who have received a one day training session.

Indicate that the test battery appears to be about as effective as the use of PBTs in improving the BAC distribution of those arrested (e.g., a reduction of false positives).

Suggest that the gaze nystagmus test is the most powerful of the three if only one is used, and that the combination of gaze nystagmus and walk and turn offers the most potential for discriminating between those above and below .10% BAC.

VII. References

1. Beitel, G., Sharp, M., and Glauz, W., <u>Probability of Arrest While Driving Under the Influence,</u> Journal of Studies on Alcohol, Vol. 36, No. 1, pp 109-116, January 1975.

2. Burns, M. and Moskowitz, H., <u>Psychophysical Tests for DWI Arrest,</u> U.S. Department of Transportation Report No. DOT-HS-802-424, June 1977.

3. Tharp, V., Burns, M., and Moskowitz, H., <u>Development and Field Test of Psychophysical Tests for DWI Arrest,</u> U.S. Department of Transportation Report No. DOT-HS-805-864, March 1981.