

FIELD SOBRIETY TESTS: ARE THEY DESIGNED FOR FAILURE?'

SPURGEON COLE AND RONALD H. NOWACZYK

Clemson University

Summary--Field sobriety tests have been used by law enforcement officers to identify alcohol-impaired drivers. Yet in 1981 Tharp, Burns, and Moskowitz found that 32% of individuals in a laboratory setting who were judged to have an alcohol level above the legal limit actually were below the level. In this study, two groups of seven law enforcement officers each viewed videotapes, of 21 sober individuals performing a variety of field sobriety tests or normal-abilities tests, e.g., reciting one's address and phone number or walking in a normal manner. Officers judged a significantly larger number of the individuals as impaired when they performed the field sobriety tests than when they performed the normal-abilities tests. The need to reevaluate the predictive validity of field sobriety tests is discussed.

Field sobriety tests have been used throughout this century by police officers to help them assess whether an individual is too impaired to drive an automobile. A classic paper by Bjerver and Goldberg, (1951) examined the relationship between performance on the field sobriety test and driving. Over the past two decades the National Highway Transportation Safety Administration (NHTSA) has funded several studies to examine the effectiveness of field sobriety tests in predicting a person's level of intoxication and driving impairment (e.g., Anderson, Schweitz, & Snyder, 1983; Burns & Moskowitz, 1977; Tharp, Burns, & Moskowitz, 1981).

In a 1977 report, Burns and Moskowitz examined a number of different tests commonly used by officers. Based on the results from a laboratory study, they recommended three tests, the Horizontal Gaze Nystagmus (HGN) test, the walk-and-turn test, and the one leg stand test for further research. The HGN measures the angle of gaze at the onset of jerking movement which can be influenced by alcohol consumption as well as other physiological factors. The other two tests require dividing attention among mental and physical tasks. Briefly, the walk-and-turn test requires a person to stand on a line in a heel-to-toe position while listening to instructions and then to take nine steps in a heel-to-toe fashion, pivot, and take nine more steps along a straight line. The one-leg stand requires an individual to stand with arms at the side and extend one foot six inches off the ground and maintain that position while counting for 30 seconds without extending the arms or losing balance. (For complete instructions see "DWI Detection and

Divided Attention Field Sobriety Testing" by NHTSA, 1987.) Although these tests seemed to hold the most promise, the authors reported that false alarms are a concern. In the 1977 study, 47 percent of the subjects who would have been arrested based on test performance actually had a blood alcohol concentration (BAC) lower than .10 percent, the decision level used by officers.

A 1981 report by Tharp, *et al* employed the three previously mentioned tests in another laboratory study. The error rate improved somewhat; 32 percent of the participants judged to have BACs greater than .10 actually had BACs lower than .10, the decision point used in many states for assuming driving impairment. Reliability coefficients for these tests, however, were often below accepted levels for standardized clinical tests. Reliable tests have coefficients of approximately .85 or higher (Rosenthal & Rosnow, 1991). Test-retest reliability coefficients for the field sobriety tests ranged from .61 to .72 for individual tests and .77 for the total test score for 77 individuals who were dosed to the same BAC level on two occasions. Interrater reliability coefficients, based on having different officers score performance on each occasion, were even lower, ranging from .34 to .60 with .57 as an over-all test score.

Problems in scoring can be attributed, in part, to the lack of standardization across many of the field sobriety test studies. In addition, a few miscues in performance can result in an individual being scored as impaired (Anderson, *et al.*, 1983). For example, a person is viewed as impaired for missing two of nine points on the walk-and-turn test or two of five points on the one-leg stand test. The stringent scoring criteria as well as potential subjectivity in determining whether a point should be awarded may account for accuracy rates that vary from 72 to 96 percent among police agencies using these tests in the Anderson, *et al.* study. The fact that these tests are largely unfamiliar to most people and not well practiced may make it more difficult for people to perform them. As few as two miscues in performance can result in an individual being classified as impaired because of alcohol consumption when the problem may actually be the result of their unfamiliarity with the test.

This study tested the hypothesis that sober individuals will find the field sobriety tests difficult to perform and, as a result, will be judged to be impaired by officers viewing their performance. Individuals who were completely sober were asked to perform several field sobriety tests and several "normal-abilities" tests which should be well known to individuals. These latter tests included answering personal data questions, such as stating one's address and phone number, as well as walking in a normal manner. Performance on the field sobriety tests and normal-abilities tests was videotaped. Law enforcement officers were asked to view these tapes and deter-

Requests for reprints can be sent to either author at the Department of Psychology, Clemson University, Clemson, SC 29634. The authors thank Ronnie Cole for his assistance in the completion of this study and Jack Davenport for his comments on an earlier draft of this manuscript.

mine if these individuals were impaired ("too drunk to drive"). If the field sobriety tests are difficult to perform under normal circumstances, then we can expect officers to judge incorrectly individuals as being impaired on the basis of the field sobriety test performance as compared with scores on the normal-abilities tests.

METHOD

Subjects and Design

Fourteen police officers from the local municipality or county sheriff's office rated the performance of 21 individuals who had completed the field sobriety and normal-abilities tests. These officers, with 1 to 17 years of law enforcement experience ($M = 11.7$ yr.) were volunteers who were certified by the South Carolina Academy for Police Officers which is a state requirement. As part of this certification requirement they had completed the state DUI training program and have had field experience with DUI detection. All officers were assigned to duties in the field.

Ten males, seven white and three African-American, and eleven white females served as participants. They were recruited from local businesses. The owners of these businesses were asked if they had any employees who were willing to volunteer to serve in an experiment involving psychomotor tasks. Participants were currently employed, between 21 and 55 years of age, and not overweight, and had no known physical disabilities.

All individuals and officers were paid for their participation. The individuals performed both field sobriety tests as well as normal-abilities tests. Half of the officers were randomly assigned to each condition in which they viewed performance on either the field sobriety or normal-abilities tests.

Tests Performed

Prior to the administration of the tests, each participant was administered the Datamaster breathalyzer test. All participants had a BAC level of .00. Each participant performed six field sobriety tests and four normal-abilities tests in the same order in an indoor setting. The field sobriety tests included the walk-and-turn test, alphabet recitation, one-leg stand, a one-leg stand while tilting backward with the eyes closed and touching the nose, a one-leg stand with counting, and a one-leg extension test. These tests were selected after interviewing a number of officers concerning tests they used in the field. None of these officers served in this study. The Horizontal Gaze Nystagmus test was not included because it requires officers individually to monitor the participants' eye movements which would have been difficult to videotape in a controlled fashion. It is also not included in the 1987 NHTSA self-instructional guide (NHTSA, 1987). The four normal-abilities tests included counting from 1 to 10, reciting one's Social Security number, driver's license number or date of birth, recit-

ing one's home address and phone number, and walking in a normal manner, turning around, and walking back to the starting point. These tests were selected by the experimenters to sample motor and cognitive activities that are commonly performed by most individuals.

Standard instructions for each test were read by the experimenter. Participants were told that they would perform a number of motor-coordination tasks that would last approximately 30 minutes. These instructions were based on those used by law enforcement in South Carolina and followed NHTSA guidelines. If participants had questions regarding the instructions, the experimenter reread the appropriate section. The reading of instructions was included on the videotape. The tests were performed indoors in a meeting room where distractions were minimal. A 7.62-cm wide strip of tape was placed on the floor for the walk-and-turn test as per NHTSA requirements.

Procedure

Each officer watched a videotape of the 21 individuals performing one of the two sets of tests. The order of performance of the individuals was the same for both the field sobriety tests and normal-abilities tests. The officers were provided with sheets of paper listing the participants by number. The officers were allowed to take notes and were asked "Do you feel, as a law enforcement officer, that the following subjects, based on field sobriety tests performed on videotape, have had too much to drink to drive.

Their responses, either "yes" or "no," were recorded for each individual. The decision was recorded by the officer immediately, after viewing the individual's performance and prior to viewing the next individual's performance. Each officer participated in individual sessions.

RESULTS

The proportion of officers who decided that an individual had "too much to drink" was recorded for each individual separately for the field sobriety and normal-abilities tests. There was a significant difference as a function of test ($t_{29} = 4.38$, $p < .01$). Forty-six percent of the officers' decisions were that an individual had "too much to drink" from viewing the field sobriety tests. Fifteen percent of the decisions from the normal-abilities tests were that a person had "too much to drink."

Differences among individuals were apparent. Only three individuals were rated as "unimpaired" by all officers on both the field sobriety and normal-abilities tests. One individual's performance was rated as showing he had had "too much to drink" based more on the normal-abilities tests (by three officers) than on the field sobriety tests (none of the officers). Five individuals were rated as having had "too much to drink" by all the officers who viewed the field sobriety tests. One other individual was rated as having had "too much to drink" by all but one officer. Of these six individuals

only one was rated as "impaired" by as many as four of the officers who saw the same individuals performing the normal-abilities tests. Four of these individuals were rated as having had "too much to drink" by two or fewer of the officers viewing the normal-abilities tests.

Discussion

The data indicate that judgments of impairment are influenced by the type of test performed. An individual was more likely to be judged as impaired on the basis of field sobriety test performance than on performance of the normal-abilities tests. Even without alcohol, the number of errors made by individuals performing the field sobriety tests was sufficient for officers to judge that the individuals had had too much to drink. These findings are consistent with other studies reporting sizable percentages of individuals judged to be impaired when they were not (Burns & Moskowitz, 1977; Tharp, *et al.*, 1981).

While training of officers, standardization of test instructions, administration, and scoring may reduce the number of incorrect classifications, the major obstacle may be the field sobriety tests. The fact that these tests require unfamiliar and unpracticed motor sequences may put an individual at a disadvantage when performing them. To the law enforcement officer who has demonstrated the tests many times, the motor sequences may seem easy and straightforward. It may also be that to the casual observer that the tests are easy to perform. Yet, when an untrained individual actually performs the test, then the difficulty of performing the tests at an acceptable level may become evident.

The reliance on field sobriety test performance by law enforcement officers in their decision to arrest or not and by juries in their decision whether to convict a person of driving under the influence underscores the need to examine field sobriety tests critically. The tests should discriminate between the two populations of individuals who are impaired and those who are not. Ideally, the tests should separate the two populations, that is, increase d , the mean difference between the two populations. The tests, however, may be doing nothing more than adjusting the officer's β , or criterion measure, downward.

These tests must be held to the same standards the scientific community would expect of any reliable and valid test of behavior. This study brings the validity of field sobriety tests into question. If law enforcement officials and the courts wish to continue to use field sobriety tests as evidence of driving impairment, then further study needs to be conducted addressing the direct relationship of performance on these and other tests with driving. To date, research has concentrated on the relationship between test performance and BAC and officers' perceptions of impairment. This study indicates that these perceptions may be faulty.

REFERENCES

- Anderson, T. E., Schweitz, M. B. (1983) M. B. (1983) Field evaluation of a behavioral battery for DWI. Final Report, DOT-HS-806476.**
- Bjerver, K. & Goldberg, L. (1951) Effect of alcohol ingestion on driving ability: results of practical road tests and laboratory experiments. Quarterly Journal of Studies on Alcohol, 11, 1-30.**
- Burns, M., & Moskowitz, H. (1977) Psychophysical tests for DWI arrest. Final Report, DOT-HS-802-424, NHTSA.**
- NHTSA. (1987) DWI Detection and Divided Attention Field Sobriety Testing. Final Report, DOT-HS-807-186.**
- Rosenthal, R., & Rosnow, R. L. (1991) *Essentials of behavioral research methods and data analysis*. (2nd ed.) New York: McGraw-Hill.**
- Tharp, V., Burns, M., & Moskowitz, H. (1981) Development and field test of psychophysical tests for DWI arrests. Final Report, DOT-HS-805-864, NH'ISA.**
- Accepted May 23, 1994.*