OVERVIEW OF STANDARDIZED FIELD SOBRIETY TESTING

RESEARCH AND DEVELOPMENT

Student Note:

The information contained in this overview is a brief summary of the research and development of SFST and does not include the complete research as contained in NHTSA Report number DOT HS-802 424.

1. First Phase: The Developmental Research

A. What were the research objectives?

- To evaluate currently used physical coordination tests to determine their relationship to intoxication and driving impairment.
- To develop more sensitive tests that would provide more reliable evidence of impairment.
- To standardize the tests and observation.

B. Who conducted the research?

Southern California Research Institute (SCRI)

The final report:

> Burns, Marcelline and Moskowitz, Herbert
> Psychophysical Tests for DWI; June, 1977
> NHTSA Report Number DOT HS-802 424
> (available from National Technical Information Service, Springfield, Virginia 22161)

C. Who were the test subjects?

They were 238 volunteers, of whom 168 were males and 70 females. They were paid $3.00 per hour, and they each participated in one testing session.

The volunteers were interviewed by SCRI staff, and on the basis of the interview they were classified as either light, moderate or heavy drinkers. They were randomly assigned to "target BAC" levels appropriate to their classifications. The following shows the distribution of BACs achieved by volunteers:

	Light Drinkers	Moderate Drinkers	Heavy Drinkers	Totals
No Alcohol (0.00%)	26	27	26	79
Approximately 0.05%	36	16	3	55
Approximately 0.075%	--	6	7	13
Approximately 0.10%	--	37	13	50
Approximately 0.15%	--	--	41	41

D. Who tested the subjects?

Ten police officers, representing four agencies in the vicinity of Los Angeles, did all of the testing. Each officer examined an average of 23-24 volunteers. While the officer was conducting the examinations, a member of the SCRI staff observed the examinations.

NOTE: Neither the volunteer nor the officer nor the observer knew the volunteer's BAC. Separate members of the SCRI staff handled the dosing and breath testing of volunteers.

E. What tests were administered?

Each volunteer was subjected to six tests:

- o One-Leg Stand
- o Finger-To-Nose
- o Finger Count
- o Walk-and-Turn
- o Tracing (a paper-and-pencil exercise)
- o Nystagmus (called "alcohol gaze nystagmus" in the final report)

Each officer was given one day's training in the administration and scoring of these tests prior to conducting the experiment. NOTE: Only two of the ten officers had any prior experience with nystagmus.

F. In general, how were the tests "scored"?

Each of the six tests were "scored" on a scale from 0 to 10; for the nystagmus test, each eye was "scored" independently, so that a subject's total nystagmus "score" could range from 0 to 20.

The higher the "score", the more impaired the subject appeared to be.

Whenever a volunteer was tested, the officer administering the test and the SCRI researcher observing the test independently scored the subject's performance.

G. What were the nystagmus administration and "scoring" procedures?

The volunteer was seated, with chin in a chin rest, and faced a small light bulb mounted on a swing arm that could be moved to precise angles on either side.

The volunteer was instructed to cover the left eye and follow the movement of the light bulb with the right eye. The officer slowly moved the swing arm to the 30-degree mark, and left it there for several seconds, while observing the volunteer's eye for jerking. "Points" were scored as follows:

no jerking	0 point
minimal jerking	2 points
moderate jerking	3 points
distinct, easily observed jerking	5 points

Next, the officer slowly moved the swing arm to the 40-degree mark, and left it there to observe the eye once again. The same scoring system was used. Then, the score for the right eye was determined by adding the scores at the 30-degree and 40-degree marks. For example, if the eye showed minimal jerking at 30 degrees (2 points) but moderate jerking at 40 degrees (3 points), the score for the eye would be 5 points.

Finally, the volunteer was instructed to uncover the left eye and cover the right eye, and the entire procedure was repeated to determine the left eye's "score". NOTE: The scores for the two eyes often were different, by a point or two.

H. What were the administration and "scoring" procedures for Walk and Turn?

The volunteer was told to stand facing the examiner (not in a heel-to-toe posture) and to "watch what I do so you will be able to do it the same way. I want you to put one foot here on the line, and then take exactly 9 steps along the line, touching your heel to your toe each step."

(The examiner then demonstrated the heel-to-toe step.)

"Then, turn and take 9 steps back along the line, touching heel-toe. (NOTE: Apparently the examiner did not demonstrate the turn.) Do you understand? Come here to the line and begin."

The officer and observer independently "scored" the volunteer's performance, using the following scheme:

no problem	0 point
falls; won't attempt test, or discontinues test	10 points
slow or minor problem in performing test	1-4 points (examiner's judgment)

Or, the examiner could assign 1 or 2 points for each of the following clues (up to a maximum of 10 points, total, for the test):

- loses balance while walking
- loses balance while turning
- cannot stay on line
- excessive movement of arms and/or body to maintain balance
- does not touch heel-toe
- incorrect number of steps
- stops to steady self
- requires repeat demonstration

I. What were the administration and "scoring" procedures for One-leg stand?

The volunteer was told to "watch what I do but don't begin until I tell you. Stand with your feet together, arms at your side, and hold one leg straight forward, like this."

(At this point, the examiner demonstrated the one-legged stance, holding a foot 8-12 inches off the floor.)

"Do you understand? Ready? Begin. Don't put your foot down until I tell you to."

NOTE: The subject was not required to count aloud for 30 seconds. Instead, the examiner simply terminated the test after 30 seconds.

The officer and the observer independently "scored" the volunteer's performance, using the following scheme:

no problem	0 point
slightly unsteady	2 points
moderately unsteady	4 points
extremely unsteady	6 points

And, 1 point was added for each of the following, if observed:

- o required a repeat of the instructions
- o put the foot down
- o used arms for balance

If the volunteer fell, or made no attempt to perform the test, or discontinued the test, they "scored" 10 points.

J. What did the researchers learn?

The researchers analyzed their data and found that, using the scores from all six tests, they could correctly classify a volunteer's BAC as being either above or below 0.10% about 83 percent of the time.

Further, the researchers found that this same level of reliability could be achieved just by considering the scores on nystagmus, walk and turn, and one-leg stand. In other words, those three tests constituted an 83% reliability battery for distinguishing BACs of 0.10% or more from BACs below 0.10%.

What about the 17% of volunteers whose BACs were misclassified? How did the researchers account for them?

First, half of the volunteers who were misclassified had BACs between 0.08% and 0.12%, a "borderline" range in which it can be difficult to distinguish slight differences in impairment. Secondly, almost all of the remaining misclassified volunteers were either light drinkers with BACs of at least 0.05% (who may well have appeared and been very impaired at that level), or heavy drinkers with BACs below 0.15% (whose experience with alcohol may have helped them mask the signs of impairment).

K. What was the overall conclusion?

The three-test battery made up of nystagmus, walk and turn, and one-leg stand clearly appeared to offer a very reliable field sobriety testing procedure. But these tests were not yet standardized in their final form. Standardization was achieved in the next phase of research.

2. The Second Phase: Initial Validation Research

A. What were the research objectives?

- To complete the development and validation of the sobriety test battery.
- To assess in the field the battery's feasibility, and its effectiveness for estimating BAC and facilitating identification of persons with BACs above 0.10%.

B. Who conducted the research?

Southern California Research Institute (SCRI)

The final report:

> Tharp, V., Burns, M. and Moskowitz, H. Development and Field Test of Psychophysical Tests for DWI Arrest, March, 1981, NHTSA Report Number DOT HS-805 864 (available from NTIS, Springfield, Virginia 22161)

C. Who were the test subjects?

During the first (laboratory) portion of this research effort, the test subjects were 296 volunteers, of whom 202 were males and 94 females. In the second (field) portion, the "subjects" were 3,128 drivers stopped by participating police officers for traffic law violations and other routine causes. Of these, the officers at least initially suspected 396 might be under the influence of alcohol or other drugs; 215 ultimately were arrested for DWI.

The 296 laboratory subjects each participated in at least one testing session. And, 145 of them returned for a second session, for a total of 441 subject-days of testing. The following table shows the distribution of these subjects by drinker classification and "target BAC"; the numbers in parenthesis refer to the subjects who returned for a second session.

	Light Drinkers	Moderate Drinkers	Heavy Drinkers	Totals
No Alcohol (0.00%)	30 (18)	32 (16)	35 (16)	97 (50)
Approximately (0.05%)	33 (15)	33 (16)	36 (17)	102 (48)
Approximately (0.11%)	--	30 (15)	34 (14)	64 (29)
Approximately (0.15%)	--	-- 33 (18)	33 (18)	

D. Who tested the subjects?

For the laboratory portion of the study, ten police officers from three agencies in the metropolitan Los Angeles area did the testing. Each officer examined an average of 44 subjects (including returnees). While the officer conducted the examinations, a member of the SCRI staff observed. Neither the volunteer, nor the officer nor the observer knew the volunteer's BAC.

For the field portion of the study, participating officers were drawn from four stations of the Los Angeles County Sheriff's Office. They included a group called the "experimentals" (who received training in the SFSTs), and a group of "controls" (who were not trained until the final stage of the study). Both groups were instructed to complete data forms for all of their traffic stops during the study period. In addition, SCRI researchers periodically rode with every officer to monitor their performance.

E. What tests were administered?

In both the laboratory and field portion of the study, participating officers (except the "controls") administered Horizontal Gaze Nystagmus, Walk and Turn, and One-Leg Stand. Some of the officers had some prior experience with these tests, but all received one-half day's training in test administration and scoring.

In addition to recording subjects' performance on the SFSTs, the officers attempted to estimate each subject's BAC.

F. How did the officers do in their estimation of subjects' BAC?

In the laboratory portion of the study, the average absolute value in the difference between officers' estimates and subjects' actual BACs (as measured on a breath testing instrument) was 0.03%. The error in the officers' estimates appeared to be random, i.e., their estimates were high about half the time and low about half the time. It should be observed that a laboratory study provides a relatively "easy" context in which to estimate BACs. All participants know (or quickly learn) that the research team will not expose the subjects to very elevated levels (e.g., 0.20% or more), and since the study design is based on fairly precise "target BACs", the subjects tend to "cluster" in the BACs they actually achieve. In other words, it would not be too difficult to make a fairly good educated guess of a subject's BAC if the officer has a reasonable amount of experience in DWI enforcement. Despite the favorable context, the officers' estimates were off by more than 0.03% about half the time.

In the study's field portion, the researchers concluded that most of the officers' estimates of subjects' BACs were invalid. Apparently, most of the participating officers filled out their data forms at the end of their shift, when they already knew the BACs of most arrestees.

G. What were the nystagmus administration and "scoring" procedures?

In the laboratory portion, two kinds of nystagmus measurements were made on each subject. First, the officer examined the subject to estimate the angle of onset, check for lack of smooth pursuit, and, check for distinct jerking at maximum lateral deviation. These checks were performed in both eyes. Second, the subject was seated at the light bulb/swing arm device used in the previous study, and a measurement of the angle of onset was obtained for each eye. In their previous research, and in their review of studies conducted by other researchers, the SCRI staff found evidence that "a strong correlation exists between the BAC and the angle of onset ..." They found that the mathematical expressions of the correlation are slightly different for the left and right eyes, but in either eye an angle of 41 degrees would correspond to a BAC of about 0.10%. They wanted to learn whether officers could estimate onset angles with reasonable precision, and whether the estimate can accurately distinguish subjects above 0.10% from those below that level.

The SCRI researchers did not report the actual data that would compare the officers' onset angle estimates with the swing arm device measurements of onset angle. Instead, they furnished a list of Pearson Product Moment Correlation Coefficients, for each officer and observer, that express how each officer's estimates "track" with the device measurements. A bit of explanation is needed in order to understand these coefficients.

In general terms, a correlation coefficient expresses the "closeness" of two sets of data. If a change in the value of one item is always associated with a systematic change in the value of the other item, then we can say that the two items are closely correlated. For example, in the summer months, there is probably a pretty close correlation between the highest daytime temperature and the number of people visiting beaches: the higher the temperature (i.e., the hotter it gets), the more people you'll find at beaches (trying to cool down). But if a change in one variable has nothing to do with changes in the other item, then we say that the two items are uncorrelated. For example, the number of people visiting beaches in America on any given day probably has nothing to do with the number of loaves of bread sold in Russia on that same day. Some days, lots of bread will get sold in Russia, and lots of Americans will go swimming. But other days, just as many Russians will buy bread, but quite a different number of Americans will be at the beach. The two items just aren't related. Another common situation occurs when two items <u>are</u> related, but the relationship is not <u>exact</u>. For example, the number of <u>runs</u> a baseball team scores in a game probably is related to the number of <u>hits</u> the team makes in the game. In other words, the more hits you get, the more likely you are to score runs. But this relationship is far from perfect. It is quite possible to get very few hits and still score lots of runs, if the other team makes lots of errors or gives up lots of walks. Runs and hits in a game probably are correlated, but the correlation may be <u>weak</u>.

The correlation coefficient gives an indication of the strength or weakness in the relationship between two items. The highest absolute value that the correlation coefficient can have is 1.00, and that occurs when the two items are <u>perfectly</u> correlated. That would mean that, if you know the value of one item you could exactly predict the value of the other item. The lowest absolute value of the correlation coefficient is 0. That occurs when the two items have absolutely nothing in common, i.e., when knowledge of the value of one is of no help at all in predicting the value of the other.

It is important to understand that two items could have a very high correlation without having <u>equal</u> values. Consider the comparison between an officer's onset angle estimations and the device-measured angles. If an officer <u>consistently</u> underestimated the device's angle by 10 degrees, we wouldn't think that the officer was very accurate. That is, if the officer said "35" when the device indicated "45", and said "40" when it indicated "50", and so on, we would consider those to be bad estimates. <u>But the correlation between the officer's estimates and the device's would be perfect</u>, and the correlation coefficient would be 1.00, simply because the relationship between the two variable would be perfectly predictable.

In reporting only the correlation coefficients for the officers' estimates the SCRI researchers aren't describing the officers' <u>accuracy</u>, but only are indicating whether there is some systematic relationship between the measured angles and each officer's estimates of them.

With all that preamble now accomplished, the correlation coefficients for the ten officers' angle estimates ranged from a low of 0.234 to a high of 0.719. Most of these correlations (at least) probably are statistically significant (although the report does not state that), but in practical terms the correlations would be considered weak to moderate. This can be quantified when the correlation coefficient is squared (i.e., multiplied by itself), the resulting number expresses the percentage of variability in one item that can be related to variability in the other item. In loose terms, it tells us how useful one item is in predicting the value of the other. For example, suppose the correlation coefficient for two items were 0.500. The square of that would be 0.250. That would mean that 25% of the variability in one item could be related to the variability of the other, or that one item would be about 25% useful in predicting the other.

The best of the ten officers had a correlation coefficient for angle estimations of 0.719. The square of that is .517. That officer's estimates are about 50% useful in predicting the "true" onset angle. The worst-estimating officer had a coefficient of 0.234, which means the estimates are about 5% useful. The average correlation coefficient for the ten officers was 0.475, indicating an average utility of a bit less than 23%.

Of course, the ability of officers to estimate onset angle is only part of the story. We also have to consider how well the "true" onset angle can predict BAC. The SCRI researchers report two different correlation coefficients for onset versus BAC, one for the left eye (absolute value of 0.780) and one for the right (absolute value of 0.740). If the higher value is accepted, then the device-measured onset angle is about 60% useful in predicting BAC.

These are not encouraging words for anyone who would claim the ability to use horizontal gaze nystagmus to "predict" BAC. The so-called "true" onset angle is only about 60% useful in predicting BAC. The average officer's estimates are less than 25% useful in predicting onset angle, and even this says nothing about any systematic inaccuracy that may exist in the officer's estimates. At best, one can expect only a 25% chance of reaching something that has a 60% chance of being useful, or overall a 15% chance of getting to anything at all. Given this, it is not surprising that these officers were off in their estimates of subjects' BACs by an average of 0.03%, despite the favorable estimation conditions of a controlled drinking experiment.

In both the laboratory and field portions of this study, officers were instructed to record the following nystagmus data, for each eye:

- Whether onset occurred within 45 degrees, with at least 10% of the white of the eye showing;
- The estimated angle of onset;
- Whether the eye was unable to follow smoothly;
- Whether the nystagmus at maximum deviation was absent, minimal, moderate or heavy.

One "point" was "scored" for each eye if onset occurred within 45 degrees; if the eye was unable to follow smoothly; and, if the nystagmus at maximum deviation was moderate or heavy.

H. What were the administration and "scoring" procedures for Walk and Turn?

Based on a review of previous research, the SCRI staff decided to modify the Walk and Turn test to incorporate a divided attention feature. Thus, the subject was instructed at the outset to "assume a heel-to-toe position on the line with your arms at your sides." The officer gave no further instructions until the subject assumed the proper stance. Then, the rest of the instructions were issued, in the same manner that they were given during the previous phase of research.

Walk and Turn "scoring" procedures also were modified, and they were slightly different for the laboratory versus field portions of this study. In the laboratory tests, officers and observers were told to "score" one "point" for each of the following behaviors:

- cannot keep balance while listening to instructions
- starts before instructions are finished
* keeps balance but does not remember instructions
- stops while walking to steady self
- does not touch heel-to-toe while walking
- loses balance while walking (i.e., steps off line)
- uses arms for balance
- loses balance while turning
- incorrect number of steps

If the laboratory subject was "unable to do the test", the officers and observers were instructed to "score" ten "points".

For the field portion of the study, the item marked above with an "asterisk" (keeps balance but does not remember instructions) was dropped, and nine "points" were given for being unable to perform the test. Thus, by the time the field study began, administration and "scoring" procedures for Walk and Turn had evolved to their present state of eight "points".

I. What were the administration and "scoring" procedures for One-Leg Stand?

SCRI researchers decided to add a divided attention feature to this test as well. The subject now was to be instructed to count aloud, "one thousand and one, one thousand and two, . . . , one thousand and thirty". Also, the instructions were modified to call for raising the foot about six inches off the ground, rather than the 8-12 inches specified during the previous research phase.

One-Leg Stand "scoring" differed slightly from the laboratory to the field portions of this study. Laboratory subjects were assessed one "point" for each of the following behaviors:

- Swaying while balancing
- Uses arms to balance

* o Slightly unsteady
o Quite unsteady
* o Starts before instructions are finished
o Puts foot down

If a laboratory subject was unable to do the test or discontinued the test, they were to be assessed seven "points".

By the beginning of the field study, SCRI researchers had dropped the two items marked with "asterisks", and were assessing five "points" for being unable to perform. Thus, One-Leg Stand had evolved very nearly to its present state of four "points". Subsequently, NHTSA staff recognized that the scoring factor "quite unsteady" was subjective; based on a re-analysis of the SCRI data, that factor was changed to "hops".

J. What did the researchers learn?

(1) The Laboratory Phase

Results of the laboratory study demonstrated that the battery of three tests could be used reliably to distinguish subjects with BACs of 0.10% or more from those with lower BACs. Collectively, the ten officers and two observers were correct in classifying subjects' BACs (above or below 0.10%) about 82% of the time. Subsequent to publication of the SCRI report, NHTSA re-analyzed the laboratory test data and found that the nystagmus test, by itself, could have produced 77% accurate classifications. Similarly, Walk and Turn was capable of 68% unaided accuracy, and One-Leg Stand of 65%. NHTSA also found that it would be possible to combine the results of nystagmus and Walk and Turn in a "decision matrix", and achieve 80% accuracy.

(2) The Field Phase

SCRI reported a number of problems that plagued the field study, chief among which was a lack of consistency by participating officers in submitting data forms. SCRI concluded that the field test data would not support in-depth statistical analysis, but nevertheless disclosed some favorable trends:

- o after training on the test battery, officers tended to make more DWI arrests; and,
- o trained officers were more accurate in identifying suspects whose BACs are above 0.10%.

The overall conclusion of this study was that the test battery works well. But it remained necessary to conduct a rigorous field test.

3. The Third Phase: Large Scale Field Validation

A. What were the research objectives?

- o To develop standardized, practical and effective procedures for police officers to use in reaching arrest/no arrest decisions;

- To test the feasibility of the procedures in operational conditions; and,

- To secure data to determine if the tests will discriminate as well in the field as in the laboratory.

In support of the first of the objectives, the NHTSA research staff began by re-analyzing the SCRI data with a view toward systematizing the administrative and "scoring" procedures for the three tests. The intent was to ensure that the tests would be quick and easy to use; that they could each be used independently of one another, i.e., if the officer elected to use only one or two of the tests; and, that they would maximize the detection of drivers at BACs of 0.10% or more while minimizing the continued investigation of persons below that level.

Essentially, the current administrative and "scoring" procedures, and "scoring" criteria, for the three tests emerged from this re-analysis.

B. Who conducted the research?

The National Highway Traffic Safety Administration (NHTSA)

The final report:

Anderson, T., Schweitz, R., and Snyder, M.
Field Evaluation of a Behavioral Test Battery for DWI
September, 1983, NHTSA Report Number DOT HS-806 475
(available from NTIS, Springfield, Virginia 22161)

C. Who were the test subjects?

They were 1,506 drivers stopped for suspicion of DWI during a three-month period during late 1982/early 1983. Of these, approximately 80% were examined using all three tests.

D. Who tested the subjects?

Police officers representing four large agencies in the eastern portion of the country did the testing. All participating officers completed a one day training session prior to the beginning of the study. The training included practice in administering the tests to volunteer drinkers.

E. What tests were administered?

The officers used the three tests that make up the Standardized Field Sobriety Testing battery. As previously noted, not all subjects were exposed to all three tests, primarily because circumstances of the stop location and/or the subject sometimes precluded use of one or two of the tests. But 89% of subjects were examined using the nystagmus test, 84% on Walk and Turn and 82% on One-Leg Stand.

F. What were the test administrative and "scoring" procedures?

The procedures followed in using and interpreting the tests were essentially those spelled out in the current NHTSA training program DWI Detection and Standardized Field Sobriety Testing (1987 Update). The tests are "standardized" in the sense that:

- o they are always administered in the same way;
- o the officer administering the tests always looks for a specific set of clues on each test; and,
- o the officer always assesses a subject's performance relative to a specific criterion for each test.

G. What are the "standardized" elements of the Horizontal Gaze Nystagmus test?

(1) Standardized Administrative Procedures

- o Hold the stimulus 12-15 inches in front of the subject's face.
- o Keep the tip of the stimulus slightly above the subject's eyes.
- o Always move the stimulus smoothly.
- o Always check for all three clues in both eyes.

NOTE: It does not matter whether you check for the three clues in one eye and then check the other eye, or check the first clue in both eyes, then the second clue in both eyes, etc. Either approach is acceptable as long as you always examine all clues in both eyes.

- o Check the clues in this sequence: lack of smooth pursuit; distinct jerking at maximum deviation; onset within 45 degrees.
- o Always check for each clue at least twice in each eye.

(2) Standardized Clues

- o Lack of smooth pursuit.
- o Distinct jerking at maximum deviation.
- o Onset of jerking within 45 degrees.

No other "clues" are recognized by NHTSA as valid indicators of horizontal gaze nystagmus. In particular, NHTSA does not support the allegation that onset angle can reliably be used to estimate BAC, and considers any such estimation to be misuse of the horizontal gaze nystagmus test.

(3) Standardized Criterion

The maximum number of clues of horizontal gaze nystagmus that a subject can exhibit is six. That would occur when all three clues are observed in both eyes. If a subject exhibits four or more clues, it should be considered evidence that they are under the influence.

H. What are the "standardized" elements of Walk and Turn?

(1) Standardized Administrative Procedures

- o Always begin by having the subject assume the heel-toe stance.
- o Verify that the subject understands that the stance is to be maintained while the instructions are given.
- o If the subject breaks away from the stance as the instructions are given, cease giving instructions until the stance is resumed.

- Tell the subject that they will be required to take 9 heel-to-toe steps down the line, to turn, and to take 9 heel-to-toe steps up the line.
- Demonstrate several heel-toe steps.
- Demonstrate the turn.
- Tell the subject to keep the arms at the sides, to watch the feet, to count the steps aloud, and not to stop walking until the test is completed.
- Ask the subject whether they understand; if not, re-explain whatever is not understood.
- Tell the subject to begin.
- If the subject staggers or stops, allow them to resume from the point of interruption; do not require the subject to start over from the beginning.

(2) Standardized Clues

- Loses balance during the instructions (i.e., breaks away from the heel-toe stance).
- Starts walking too soon.
- Stops while walking.
- Misses heel-to-toe while walking (i.e., misses by at least one-half inch).
- Raises arms from side while walking (by six inches or more).
- Steps off the line.
- Turns improperly.
- Takes the wrong number of steps.

These eight are the only validated clues of Walk and Turn. However, officers may see or hear other noteworthy evidence while the subject is performing this test, and officers should include any such observations in their reports.

Officers should note in their reports how many times each of the eight clues appears. However, for purposes of applying the standardized criterion (discussed below), a clue should be "counted" only once, even if it appears more than once.

If the subject cannot perform or complete the test, it should be considered that they have exhibited nine clues. One situation that would warrant this is if the subject steps off the line three or more times.

(3) Standardized Criterion

If a subject exhibits at least two clues on Walk and Turn, it should be considered evidence that they are under the influence.

I. What are the "standardized" elements of One-Leg Stand?

(1) Standardized Administration Procedures

- Tell the subject to stand with heels together, and arms at sides.
- Tell the subject not to start the test until you say to do so.
- Ask the subject if they understand.
- Tell the subject they will have to stand on one foot, with the other foot about six inches off the ground.
- Demonstrate the stance.
- Tell the subject to count from 1 to 30, by thousands.
- Demonstrate the count, for several seconds.
- Ask the subject whether they understand; if not, re-explain whatever is not understood.
- Tell the subject to begin.
- If the subject stops or puts the foot down, allow them to resume at the point of interruption; do not require the count to begin again at "one thousand and one".

(2) Standardized Clues

- Sways
- Puts foot down
- Hops
- Raises arms from side (six inches or more)

These are the only four validated clues of One-Leg Stand. However, officers may see or hear other noteworthy evidence while this test is being performed, and should include any such evidence in their reports.

If the subject cannot perform or complete the test, it should be considered that they have exhibited five clues. One event that would warrant this is if the subject puts the foot down three or more times.

(3) Standardized Criterion

If the subject exhibits two or more clues on One-Leg Stand, it should be considered evidence that they are under the influence. As with Walk and Turn, clues should be counted only once in applying this criterion.

J. What did the researchers learn?

The three standardized tests were found to be highly reliable in identifying subjects whose BACs were 0.10% or more. Considered independently, the nystagmus test was the most accurate of the three among subjects who exhibited four or more clues 82% had BACs of 0.10% or higher. However, the other two tests were nearly as accurate (80% for Walk and Turn, 78% for One-Leg Stand). When the nystagmus and Walk and Turn results were jointly interpreted using the decision table, 83% of the subjects were classified correctly.

The importance of this large scale field validation study deserves to be emphasized. It was the first significant assessment of the "workability" of the standardized tests under actual enforcement conditions, and it was the first time that completely objective clues and scoring criteria had been defined for the tests. The results of the study unmistakably validated the SFSTs.

But it is also necessary to emphasize one final and major point. This validation applies only when the tests are administered in the prescribed, standardized manner; and only when the standardized clues are used to assess the subject's performance; and, only when the standardized criteria are employed to interpret that performance. If any one of the standardized test elements is changed, the validity is compromised.