The National Highway Traffic Safety Administration

The Robustness of the Horizontal Gaze Nystagmus Test

NHTSA

www.nhtsa.gov

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle The Robustness of the Horizontal Gaze Nystagmus Test		5. Report Date September 2007 6. Performing Organization Code
7. Author(s) Marcelline Burns, Ph.D.		8. Performing Organization Report No.
9. Performing Organization Name and Address Southern California Research Institute 11914 W. Washington Boulevard Los Angeles, California 90066		10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTNH22-98-D-55079
12. Sponsoring Agency Name and Address U.S. Department of Transportation National Highway Traffic Safety Administration 1200 New Jersey Avenue SE. Washington, DC 20590		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code

15. Supplementary Notes

The Contracting Officer's Technical Representative for this project was Dr. James Frank.

16. Abstract

Police officers follow procedures set forth in the NHTSA/IACP curriculum when they administer the Standardized Field Sobriety Tests (SFSTs) to suspected alcohol-impaired drivers. The SFSTs include Horizontal Gaze Nystagmus (HGN) test, Walk-and-Turn (WAT) test, and One-Leg Stand (OLS) test. Courts generally accept testimony about WAT and OLS, but may not admit testimony about HGN. It has been argued that variations from standard procedures in HGN administration affect its validity and should render testimony about it inadmissible.

Three experiments examined the effects of procedural variations in administration of the HGN test. Variations in stimulus speed and elevation, and distance of the stimulus from the suspect's face were examined in a laboratory experiment. A second experiment conducted in training workshops varied the participants' positions (standing, sitting, lying down). The third experiment examined HGN in participants who have functional vision in only one eye.

The data demonstrate the validity of the HGN test with both standard and varied testing procedures. The variations did not alter the occurrence of, or the observations of, HGN.

17. Key Words	18. Distribution Statement
Nystagmus, Horizontal Gaze Nystagmus (HGN), Standardized Field Sobriety Tests (SFSTs), Driving-under-the-influence (DUI)	Document is available through the National Technical Information Service Springfield, VA 22161 and free of charge at www.nhtsa.dot.gov

19. Security Classif.(of this report)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified		

Form **DOT F 1700.7** (8-72) **Reproduction of completed page authorized**

This page intentionally left blank.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The battery of roadside tests known as the Standardized Field Sobriety Test (SFST) battery is widely used by police officers. They typically are allowed to testify in driving-under-the-influence (DUI) trials about two of the tests: Walk-and-Turn and One-Leg Stand. Their testimony about Horizontal Gaze Nystagmus (HGN), however, is frequently challenged by defense attorneys. Recently, the challenges have focused on administration procedures with arguments that any deviation from NHTSA guidelines, as laid out in training curricula, invalidates the test. To obtain data relevant to this issue, variations in HGN administration were examined in a study that included laboratory experiments and field data collection. The data for the study were records of HGN observations by SFST-trained and experienced officers, who examined alcohol-dosed participants under standardized and altered administration conditions.

An HGN examination requires a suspect to follow the movement of a stimulus with the suspect's eyes. The effects of variations in the speed of stimulus movement, in the elevation of the stimulus relative to eye level, and in the distance between the stimulus and the face were examined.

HGN examinations typically are conducted with a suspect standing, feet together and arms at the side. Occasionally, however, a suspect is unable to stand, and observations are obtained in a sitting or lying-down position. The validity of those examinations has been challenged. To examine this issue, records were obtained from the examinations with more than 900 drinking participants in training workshops assuming the three different positions.

Officers examine a suspect's eyes and score each eye separately. Questions have arisen about those observations if the individual has functional vision in only one eye. An experiment was conducted with a sample of people with monocular vision.

The principal findings are the following:

- A stimulus speed that is faster than two seconds center-to-side and side-to-center results in false negative errors.

- Holding the stimulus closer to a suspect's face (10") increases the number of HGN signs correctly observed. The gain is relatively small, however, and must be weighed against considerations of officer safety.

- In laboratory experiments, officers made no observational errors when participants' blood alcohol concentrations (BACs) were $\geq$ .10 grams per deciliter (g/dL) and very few errors when participants' BACs were $\geq$.08.

- Participants' position (standing, sitting, or lying down) had no statistically significant effects on officers' reports of HGN signs.

- HGN appears to be reduced in a non-functioning eye. If officers were to rely solely on eye signs, this reduction in HGN signs in non-functioning eyes would only increase officers’ false-negative rates and they might improperly release individuals with monocular vision. There is no evidence that HGN signs in such individuals will lead to false arrests.

It is concluded that HGN is a robust phenomenon.

This page intentionally left blank.

The Robustness of the Horizontal Gaze Nystagmus Test

I. BACKGROUND AND RATIONALE

In the 50 United States, law enforcement officers use the Standardized Field Sobriety Test battery at roadside. They are trained in accordance with NHTSA/IACP guidelines to administer, observe, and score drivers' performances on three tests (Table 1). When a trained officer asks a driver to perform the tests at roadside, that driver's performance becomes one component of the information that leads to a release or an arrest for alcohol and/or drug impairment.

Table 1.
The Standardized Field Sobriety Test Battery

The Three-Test SFST Battery
Horizontal Gaze Nystagmus (HGN)
Walk-and-Turn (WAT)
One-Leg Stand (OLS)

When an officer makes an arrest for driving-under-the-influence of alcohol or driving-under-the-influence of drugs (DUID), the officer may later testify about the evidence that led to the arrest, including the driver's performance on the SFST. Courts have generally accepted testimony about WAT and OLS tests, but testimony about HGN sometimes has not been admitted, despite the fact that of all three tests this test is the most highly correlated with BAC.

Legal challenges to the SFSTs focused largely on questions of validity and reliability (Bobo, 2003). More recent arguments have been that the tests should be inadmissible if an officer's administration of them varied from standardized procedures as set forth in the NHTSA/IACP curriculum. Specifically, it has been argued that if any detail of roadside use of the SFST departed from the NHTSA guidelines, that variation invalidates the test results.

Arguments in court about HGN evidence typically fail to distinguish between (1) essential components of test administration that both *can* and *must* be followed, and (2) procedural details that are unlikely to be critical to correct decisions about impairment. Deviations from the essential components of test administration change and possibly invalidate the tests. To illustrate, the validity of the examination hinges on complete instructions and on a suspect's understanding of those instructions. Although it is not necessary for officers to follow training manual wording verbatim, substantive changes or omissions concerning what the suspect is being asked to do are not acceptable.

Minor procedural differences occur for a variety of reasons. The environment, weather, and the suspect's level of cooperation can make adaptations advisable, even necessary. Also, officers do not have standardized instruments with which to make precise measurements in the field. These sources of variability constrained test selection during the initial development of the battery.

Table 2 provides a description of HGN signs and procedures. The term *stimulus* in the table refers to the officer's pointer, which may be a finger, a pen, a penlight, or other similar object upon which the suspect is instructed to fix his or her gaze.

Table 2.
The Three HGN Signs, the Appearance of the Eye, and the Standardized Procedures for Each Test

Sign	Appearance	Standardized Procedures
Lack of smooth pursuit (LSP)	Eye does not follow a moving stimulus smoothly.	Stimulus rate (speed of pass) is 2 seconds.
Nystagmus at maximum deviation (MAX)	With the eye gazing as far to the side as possible, jerking is distinct.	Stimulus is moved laterally to the extreme gaze possible and is held at that position for $\geq$4 seconds.
Onset of nystagmus prior to a 45^o angle of gaze (AOG)	As the eye moves to the side, jerking occurs before the eye reaches 45^o angle of gaze (AOG).	Stimulus is moved slowly to determine the AOG where jerking first occurs.

Within the standardized procedures specified in Table 2, there may be some variations in roadside test administration, but no evidence has been reported that these minor variations change either the *occurrence* of HGN signs or an officer's *observations* of them. However, because this assumption has been challenged and because the topic had not been systematically examined, this study of the effects of a set of procedural variations was conducted. The general research hypothesis of this study was that the variations do not affect the accuracy of the HGN observations and the validity of conclusions based on them.

Prior to initiation of the experiments, a comprehensive literature search was conducted using Web search engines. The objective was to determine whether the questions to be addressed by the study had been examined and published in the scientific literature on alcohol, drugs, traffic safety, and law enforcement. At the time the literature search was conducted, a large number of HGN listings were found, but most were not specific to the use of HGN as a sobriety test. Many that did address sobriety testing were not in scientific journals, having appeared instead in non-refereed DWI defense conference proceedings and newsletters. No studies were found that dealt specifically with the effects of variations in HGN administration on its accuracy and validity.[1]

Because HGN serves as a diagnostic tool for emergency room physicians, neurologists, and ophthalmologists, a computer search of the medical literature was also performed. No papers that pertained directly to the topics of this study were found.

[1] Since the project began, a study was conducted in parallel by Citek, Ball, and Rutledge (2003) which will be discussed later in the report.

II. METHOD

II.1. Design

Three experiments were conducted to evaluate the effects of several procedural variations in the administration of the HGN. They are listed and described in Table 3.

Table 3.
Study Variables – HGN Procedural Variations Examined in This Study

Experiment	Variable	Definition
I	Stimulus speed	Rate of speed at which the stimulus moves as it passes in front of a participant's eyes
	Stimulus elevation	Vertical position of the stimulus relative to the participant's eye-level gaze
	Stimulus distance	Distance of the stimulus from the participant's face
II	Participant's posture	Participant standing, sitting, or lying down during examination
III	Participant's vision	Participant having monocular versus binocular vision

The effects of differences in procedural variables could arise in two different ways:

1) In the actual occurrence of HGN signs. The argument is that the stimulus speed and position relative to the eye, the participant's posture, and the condition of having only one functional eye may affect the *occurrence* of HGN signs. It has been alleged, for example, that if the stimulus is elevated above the standard position, the resulting activation of different eye-control muscles will alter the occurrence of observable signs.

2) In the officer's perception of the signs. The stimulus speed and position, the participant's posture, and the condition of only one functional eye may affect officers' *observations* or perception of the signs. These variables have been standardized in terms of a specific speed of stimulus movement, an approximate elevation of the stimulus relative to the participant's eye level, and a range of distances between the stimulus and the suspect's face. Suspects most often are examined while in a standing position. The standardized conditions and values chosen are those that were found in laboratory study to facilitate both the officer's view of the participants' eyes, and participants' view of the stimulus.

The data obtained in three experiments were police officers' reports of their observations of HGN signs under various conditions. However, if their observations were found to differ as a function of different conditions that data alone would not have identified the origin of the differences. That is, we would still not know if the condition affected the *observations,* or if it affected the actual *occurrence* of the signs. To assist in answering that question, the laboratory examinations were videotaped to permit repeated, independent reviews of the participants' eyes during examination for the purpose of resolving questions about the origin of the differences.

The study methodology and informed consent were submitted to the Southern California Research Institute (SCRI) Institutional Review Board (IRB) for review and was approved prior to initiation of the experiments.

III. EXPERIMENTS

III.1. Experiment 1

Experiment 1 consisted of 3 mini-studies: one on the effects of variations on stimulus speed (Speed Study), one on the effects of variations on stimulus elevation (Elevation Study), and one on the effects of variations on stimulus distance from the face (Distance Study). Each mini-study used the same methodology described in section III.1.A below.

III.1.A Method: Experiment 1

There were nine participants for each mini-study resulting in a total of twenty-seven participants. Each mini-study was a repeated measures design.[2] Participants were tested under all conditions in each mini-study for four testing periods to assess the results of varying HGN stimulus administration at different BAC levels.

- In the Speed Study, participants received an HGN examination with the standard 2 second stimulus speed and a 1 second stimulus speed at 4 testing periods resulting in 8 HGN examinations per participant.
- In the Elevation Study, participants received an HGN examination with the stimulus held at the standard 2 inches above eye level, at eye level and 4 inches above eye level at 4 testing periods resulting in 12 HGN examinations per participant.
- In the Distance Study, participants received an HGN examination with the stimulus held at the standard 12" to 15" in front of the face, 10" in front of the face, and 20" in front of the face at 4 testing periods resulting in 12 HGN examinations per participant.

Police officers were recruited as examiners, and each was scheduled to participate in three successive sessions.

The research hypotheses of Experiment 1 appear in Table 4.

[2] This study was a repeated measures design which reduces error between participants due to individual differences; allows for a smaller sample size than would be necessary for a between-subjects design; and has more statistical power than a between-subjects design.

Table 4
Variables and Hypotheses of Experiment 1

Stimulus Variation	Hypotheses
Speed	The *occurrence* of a lack of smooth pursuit movement with 1-second stimulus speed = The *occurrence* of a lack of smooth pursuit movement with the standard 2-second stimulus speed.
	SFST-trained officers' *observations* of the lack of smooth pursuit movement with 1-second stimulus speed = SFST-trained officers' *observations* of the lack of smooth pursuit movement with 2-second stimulus speed.
Vertical position	The *occurrence* of HGN signs with the stimulus held either at eye level or 4 inches above eye level = The *occurrence* of HGN signs with the stimulus held 2 inches above eye level.
	SFST-trained officers' *observations* of HGN signs with the stimulus held at eye level or 4 inches above eye level = SFST-trained officers' *observations* of HGN signs with the stimulus held 2 inches above eye level.
Position in front of face	The *occurrence* of HGN signs with the stimulus held 10" or 20" in front of the participant's face = The *occurrence* of HGN signs with the stimulus held in the range 12" to 15" in front of the participant's face.
	SFST-trained officers' *observations* of HGN signs with the stimulus held 10" or 20" in front of the participant's face = SFST-trained officers' *observations* of HGN signs with the stimulus held 12" to 15" in front of the participant's face.

1. Apparatus

A Video/HGN System (EyeDynamics, Inc) was used to make video records of participants' eyes during examinations. The apparatus uses a small adjustable camera mounted in the right side of goggles that are worn by the participant. The camera transmits an image of the participant's right eye to a television monitor and VCR which the examiner used to view the right eye. The open left side of the goggles allows the participant's left eye to be viewed by the examiner.

To videotape an examination, a research assistant placed the goggles on the head of the participant, secured the goggles with adjustable Velcro straps, turned on and focused the camera, and adjusted the camera position so the eye was completely visible on the monitor. Using a wireless microphone, he then recorded date, time, participant number, test number, and experimental condition. The officer used the microphone to record the specific sign being observed as the examination proceeded, but not his or her conclusion. This recorded information for each tape segment ensured that reviewers of the videotapes would not err in identifying participants and conditions. The officers followed administrative guidelines concerning stimulus distance, height, and speed, and recorded their observations of the participant's left eye as the instrument videotaped the right eye.

2. Participants

Twenty-seven participants were recruited via Craig's List (http://losangeles.craigslist.org); a large, multi-city Web site where notices about full-time or part-time jobs are posted. The content of the posting on the Web site was as follows:

Southern California Research Institute (SCRI) needs volunteers for an alcohol study. You are eligible if you:
o Are age 21 or older. o Have a valid driver's license. o Are willing to drink alcohol. o Are available ≤ 8 hours, 10 a.m. -6 p.m. The average time of participation is expected to be 5-6 hours. o Live within the Culver City or Santa Monica area.
Qualified individuals will be paid $75 for (1) session. Call Steven at (phone number was provided in the announcement) to schedule an appointment or to leave a message w/contact information. For information about SCRI, visit www.scri.org. Thank you!

Applicants 21 or older were eligible if they were licensed drivers, moderate alcohol consumers to low-heavy alcohol consumers,[3] and if they reported current good health. They were interviewed by telephone with SCRI's standard intake questionnaire. The interviewer obtained gender, age, height, weight, and availability information. To encourage candid answers, all questions were asked while the caller was anonymous and before details of study requirements were provided.

The telephone call was terminated without the identification of ineligible individuals. If it initially appeared that the applicant would meet study criteria, the alcohol Quantity-Frequency-Variability (Q-F-V) questionnaire (Cahalan, Cisin, & Crossley, 1969) was administered to assess alcohol use. Finally, if Q-F-V responses about the frequency and amount of alcohol beverage consumption indicated moderate drinking, study requirements and conditions were described fully. People who, at that point, expressed a desire to participate were enrolled and scheduled. A participant roster appears in Appendix I.

Any of the following characteristics or conditions rendered an applicant ineligible:

- Alcohol abstainer or light drinker;[4]

[3] Due to ethical and practical concerns it was unfeasible to recruit abstainers, light drinkers and heavy drinkers.
[4] A light drinker cannot tolerate the amount of alcohol required to produce BAC's necessary for this project. In addition, a light or infrequent drinkers may show impairment at small doses (Burns & Moskowitz, 1977).

- Heavily consumes alcohol or is an alcoholic;[5]
- Current moderate-to-heavy drug use (illicit or medicinal);
- Unable or unwilling to abstain from all drug use for 48 hours prior to a session;
- Unwilling to provide urine specimen for drug screen upon request;[6]
- Injury or hospitalization within six months; and/or
- Self-reported health problem, acute or chronic.

3. Officer-Examiners

To ensure that officers who participated in the experiment were currently proficient with HGN administration, eligibility was limited to members who met all of the criteria specified in Table 5.

Table 5
Criteria for Participation for Police Officer Examiners

Category	Criteria
Traffic Officers	SFST trained with NHTSA/IACP curriculum
	Completed training more than one year ago
	Has administered SFSTs in the field ≥ 50 times
Certified Drug Recognition Experts (DRE)	Obtained certification more than one year ago
	Currently in good standing
	Currently conducts drug evaluations
	Has conducted ≥ 25 evaluations post-certification
SFST or DRE Instructors	Currently in good standing

Seven officers were recruited through a computer-based network of certified DREs (the DRE listserv) and by word of mouth. Their services are often in high demand by their agencies, and therefore the number of officers available for each session varied, as indicated in Table 6. Two officers were available for the first three sessions; three for sessions 4-6; and three for sessions 7-9. An officer roster appears in Appendix II.

[5] There are ethical considerations for providing alcohol to chemically-dependent individuals (heavy alcohol consumers).
[6] This condition was used to decrease the chance that participants in the study were drug-impaired. While no drug testing was done, participants were asked this question at the initial screening and again in person with the assumption that drug impaired individuals would not consent to a urine specimen and could be excluded from the study.

Table 6
Participation of Police Officers, by Session

Officer	Session
1	1, 2, 3
2	1, 2, 3
3	4, 5, 6
4	4, 5, 6
5	4, 5, 6, 7, 8, 9
6	7, 8, 9
7	7, 8, 9

4. Session Procedures

One session was scheduled per week for nine weeks. Participants were instructed to consume no food or stimulant beverage for four hours prior to arriving. They were transported to the SCRI facility by taxi. Upon arrival for a session at 8 a.m. to 9 a.m., they completed a questionnaire about food and caffeine intake and hours of sleep prior to the session and their answers are detailed in Table 7. The Q-F-V questionnaire was re-administered to verify responses obtained during the telephone interview.

a. Alcohol Dosing

The experiment was conducted in a double-blind procedure with neither participants nor officers aware of the amount of alcohol administered or the measured BACs. Officers did not observe the participants while they were drinking and had no contact with them prior to the first examination.

The alcohol dosing procedure is consistent with previous studies (Burns & Moskowitz, 1977; Tharp, Burns & Moskowitz, 1981). In each group of three participants, two participants were dosed to reach their highest BACs after a 30-minute drinking period followed by a 30-minute absorption period. The third participant in each group of three was given alcohol to reach an intermediate or low BAC initially and then received a booster dose of alcohol following the first examination period. This dosing scheme was developed to offset expectations the officers might have developed about participants' alcohol levels. The varied times of peak BACs and normal variations in metabolism rates served to eliminate confounding of alcohol levels with time since start of the session.

Alcohol dose amounts were based on gender, age, and body composition, and were calculated to produce expected peak BACs $\leq$ .12 at the end of the 30-minute absorption period. The alcohol beverage was 80-proof vodka and orange juice, mixed two parts orange juice with one part vodka. Three equal-volume drinks were given at 10-minute intervals with the entire amount being consumed within 30 minutes. Participants were monitored continually during the drinking and absorption periods, and breath testing with AlcoholSensor IV breath test devices began 30 minutes after drink completion. BACs were measured immediately prior to and following all examinations. When participants' measured BACs declined below .03, they were taken to their residences by taxi.

Table 7
Participants' Responses to Food, Caffeine, and Sleep Questionnaire

	Sleep		**Food**		**Caffeine**
Participant	Hours of Sleep	Hours elapsed since awakening	Food consumed	Hours elapsed since awakening	Hours elapsed since last consumption
1	8.0	3	Snack	¼	>24
2	7.5	3	Full meal	1 ½	½
3	7.0	4 ¼	Bagel	8	>24
4	9.0	1	Full meal	14	None
5	6.5	4 ½	Quesadilla	2 ½	2 ½
6	9.0	1 ½	Fruit	1	>24
7	9.0	1	Cereal	1	1
8	8.0	5	Cereal, toast, eggs	4 ½	None
9	8.0	3	Toast	3	16
10	7.0	4 ½	Donuts	1 ½	1 ½
12	8.5	2 ¾	Full meal	14	1 ¾
13	6.75	3 ½	Full meal	11	3
14	7.5	2	Celery with peanut butter	12	17
15	7	4 ½	Orange juice, egg, potatoes	3	2
16	6	4	None	N/A	16
17	7	4 ½	½ English muffin	4 ½	20 ½
18	5.5	3 ¾	Oatmeal, tea, orange juice	10 ¼	10 ¼
19	5	3	Pasta, salad, chips & salsa	16	14
20	7.5	2	Full meal	13 ½	13 ½
21	8.5	2 ¼	Full meal	13 ½	13 ½
22	8	2 ½	Sandwich, juice, grapes	11	>24
23	7	2	Chicken liver, crackers	12	11 ½
24	9	4 ¾	Waffles, meat, potatoes, rice	4 ¼	>24
25	7.5	2 ½	Eggs	11 ½	22
26	6	5 ½	Cottage cheese, yogurt	4	14 ½
27	9	4 ½	Pasta	>24	17 ½

b. HGN Examinations

Table 8 displays the three conditions for the three test variables. Prior to examinations, the officers practiced and demonstrated to the investigator's satisfaction that they could execute the conditions to which they were assigned. Participants stood in front of a large clock and officers viewed the second hand to comply with the required stimulus speed.

Successive contacts with participants were separated by one hour, as well as by examinations of two other participants. Nonetheless, as soon as a data sheet was complete, the officer deposited it in a designated file. Not having access to scores for the examinations minimized the potential for observations at one time to influence subsequent observations. One officer, one participant, and a research assistant were present in the testing area for each examination. The investigator was always in the general area and observed many of the examinations.

Table 8
Test Variables and the Stimulus Characteristics for Three Test Conditions

Stimulus elevation	Stimulus distance from face	**Stimulus speed**		
		2 sec (standard)	1 sec	
2 in (standard)	12 – 15 " (standard)	Condition 1	Condition 2	
		Stimulus elevation		
Stimulus speed	Stimulus distance from face	2" (standard)	0"	4"
2 sec (standard)	12 – 15 " (standard)	Condition 4	Condition 5	Condition 6
		Stimulus distance from face		
Stimulus elevation	Stimulus speed	12 – 15" (standard)	10"	20"
2 in (standard)	2 sec (standard)	Condition 7	Condition 8	Condition 9

III.1.B Results: Experiment 1

For all analyses, the six indicators of officer-observed HGN (three signs for each eye) yielded a scale varying from zero (no indicators) to six (all three signs in both eyes). These HGN scores were recorded for each of four exams in all conditions. Thus, for each of the variables that were examined, 9 participants provided a maximum of 36 HGN scores (e.g., 9 participants times 4 tests). Distributions of observed HGN scores were acceptable given the sample sizes for all analyses.

Planned analyses included five comparisons: (1) increased versus standard speed, (2) reduced versus standard distance, (3) increased versus standard distance, (4) reduced versus standard elevation, and (5) increased versus standard elevation. All planned analyses were preceded by a "2 x 4" within-subjects analysis of covariance (ANCOVA): two conditions by four exams with BAC as a time-varying covariate. None of the interactions were statistically significant ($p > .05$), so comparisons between conditions were averaged over the four exams for all planned analyses.

Most statistical tests assume that the null hypothesis is one of equal means between groups and to reject the null hypothesis means that the groups are different. However, the objective of this experiment was to show that there is no difference in HGN signs when stimulus speed, elevation and distance vary from the standard. Therefore, a statistical test of equivalence was the appropriate test to conduct where the null hypothesis assumes that groups are different for each condition and a rejection of the null hypothesis means that groups are similar.

Analyses of covariance through SPSS provided information about HGN means and standard errors adjusted for BAC. These values were then used in the SOLAS software, which develops confidence intervals for differences between means as well as tests of equivalences of pairs of means.

Equivalence between means for a pair of conditions were evaluated in EquivTest (Statistical Solutions) using a Schuirmann OST/TOST procedure in which one-tailed *nonequivalence* tests are applied to each of two null hypotheses: one for an upper bound of nonequivalence between means and another for a lower bound. A meaningful difference using a criterion of ± 20% (Type II error) of the HGN scores in the baseline condition was used for the Schuirmann tests. Rejection of the null hypothesis for these tests in both directions indicates that equivalence can be claimed. ***Non-rejection of either test is associated with nonequivalence.***

The mean BAC for 27 participants was .094, and varied from .044 to .143 (Figure 1, Table 9). Each participant was tested four times (Test Periods). Sixteen participants reached their peak BACs at the first test period. Eleven participants reached their peak BACs at later periods either because they absorbed the alcohol more slowly or because they were given booster doses of alcohol after the first test period.

Figure 1
Mean BACs by Test Period and Stimulus Condition

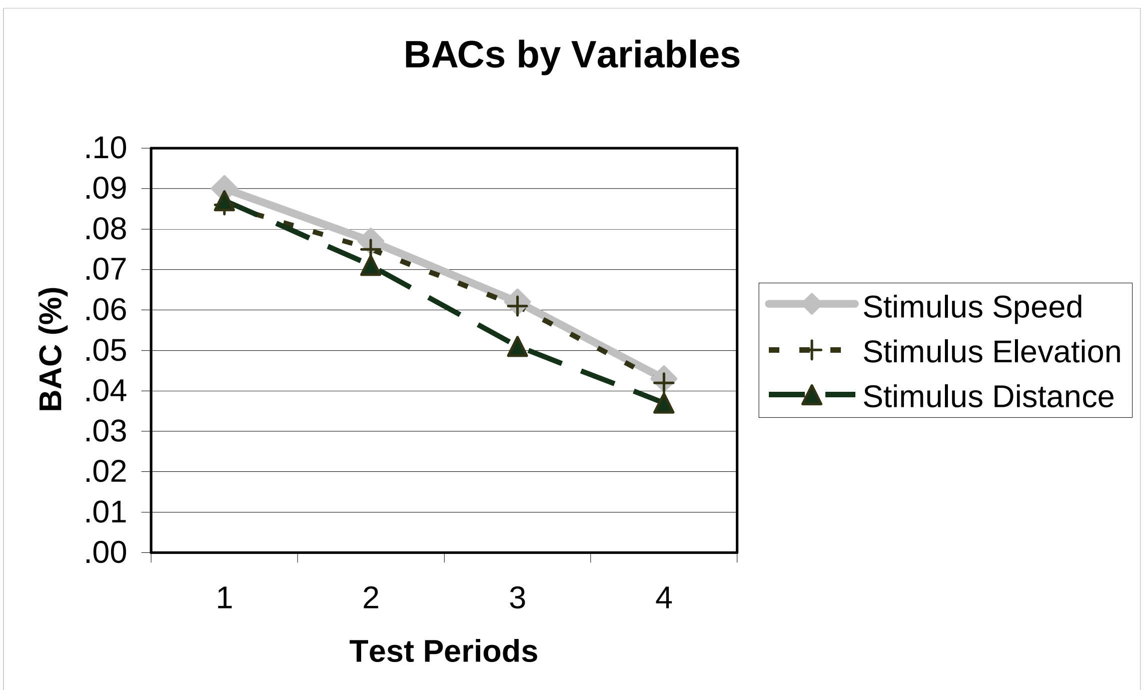


Table 9
Participants' BACs. By Test Time and Test Variable

S	Stimulus Variable	Blood Alcohol Concentration (g/dL)								
		Test Period 1		Test Period 2		Test Period 3		Test Period 4		
		Pre-Test	Post-Test	Pre-Test	Post-Test	Pre-Test	Post-Test	Pre-Test	Post-Test	Peak
1	Speed	.102	.096	.073	.067	.044	.039	.016	.011	.102
2	Speed	.087	.101	.069	.065	.049	.043	.030	.027	.101
3	Speed	.075	.064	.055	.054	.063	.050	.044	.042	.075
4	Speed	.143	.128	.127	.117	.095	.096	.076	.077	.143
5	Speed	.097	.086	.088	.084	.059	.059	.042	.039	.097
6	Speed	.064	.054	.054	.049	.073	.074	.057	.054	.074
7	Speed	.084	.089	.084	.080	.069	.068	.058	.052	.089
8	Speed	.093	.086	.063	.060	.039	.037	.019	.018	.093
9	Speed	.063	.068	.077	.072	.063	.060	.048	.044	.077
10	Elevation	.069	.071	.069	.061	.091	.088	.072	.068	.091
11	Elevation	.100	.088	.074	.075	.059	.057	.034	.031	.100
12	Elevation	.109	.102	.081	.081	.063	.064	.052	.042	.109
13	Elevation	.102	.110	.094	.093	.063	.063	.038	.036	.110
14	Elevation	.068	.071	.089	.093	.074	.072	.049	.050	.093
15	Elevation	.060	.065	.047	.047	.031	------	.019	.016	.065
16	Elevation	.079	.100	.086	.080	.065	.063	.047	.041	.100
17	Elevation	.099	.091	.080	.079	.060	.057	.038	.036	.099
18	Elevation	.092	.079	.059	.055	.047	.045	.033	.031	.092
19	Distance	.087	.070	.052	.047	.035	.035	.022	.020	.087
20	Distance	.104	.092	.076	.069	.049	.050	.027	.025	.104
21	Distance	.079	.059	.045	.046	.029	.029	.012	.011	.079
22	Distance	.070	.072	.075	.075	.057	.055	.046	.042	.075
23	Distance	.129	.119	.102	.096	.070	.071	.056	.041	.129
24	Distance	.066	.069	.106	.099	.079	.076	.059	.055	.106
25	Distance	.112	.106	.085	.081	.063	.060	.037	.033	.112
26	Distance	.091	.080	.064	.062	.051	.046	.036	.032	.091
27	Distance	.044	.039	.036	.035	.028	.025	------	------	.044
	Mean	**.088**	.084	**.074**	.071	**.058**	.057	**.041**	.037	.094
	Std. Dev.	.022	.021	.020	.019	.017	.017	.017	.016	.020
	Pre & Post Mean	.086		.073		.058		.039		
		Δ		.013		.015		.019		

1. The effects of stimulus speed

One officer examined smooth pursuit movements with a standard two-second pass of the stimulus (Condition 1). A second officer moved the stimulus quickly with a one-second pass. Although effects of variations in the speed of the stimulus were expected to be limited to smooth pursuit movements and to be reflected only in the LSP measure, the officers also scored nystagmus at the extreme lateral deviation of the eyes (MAX) and determined the angle of gaze at the onset of nystagmus (AOG). The total scores based on the three measures are summarized in Table 10, and the detailed statistical analysis appears as Appendix III. In statistical tests of nonequivalence, a first analysis compared total signs averaged over examinations at four test times. A second analysis was restricted to total signs at the first test time. Statistical equivalence was not found in either analysis, indicating that significantly more signs were reported when the stimulus movement duration was two seconds than when it was twice as fast at one second.

Table 10.
HGN Signs (Total Number) by Variations of Stimulus Speed, BACs, and Examination Period

			Stimulus Speed	
BACs (g/dL)	**Exam**	**Participant's BAC (g/dL)**	**2 sec (Standard)**	**1 sec**
≥ .100	**1**	.102	4	6
		.143	6	4
	2	.127	6	4
.05-.099	**1**	.097	4	0*
		.093	4	4
		.087	6	4
		.084	6	4
		.075	2*	3*
		.064	4	4
		.063	6	0*
	2	.088	4	2*
		.084	6	4
		.077	4	2*
		.073	6	4
		.069	6	4
		.063	4	4
		.055	2	2
		.054	4	4
	3	.095	6	4
		.073	4	4
		.069	4	4
		.063	4	0*
		.063	2*	0*
		.059	2	0*
	4	.076	4	0*
		.058	4	4
		.057	4	4
<.05	**3**	.049	4	4
		.044	4	4
		.039	4	0
	4	.048	4	0
		.044	2	2
		.042	4	0
		.030	4	2
		.019	2	4**
		.016	4**	4**

*False Negative (FN) and **False Positive (FP) relative to the ranges specified above for the various BACs

The statistically significant findings of more signs at two seconds were not completely unexpected, but still raise a number of important questions. Are the officers' observations correct; that is, are they congruent with measured BACs? How do the observations of LSP (maximum = 2), which is the component measure that is most likely to be affected by stimulus speed, relate to the total scores (maximum = 6)? Further analyses and review of videotapes of the examinations were undertaken to answer these and other questions.

"Total Score" Errors are summarized in Table 11. The criteria by which scores have been classified as correct, false negative, or false positive as defined in the SFST curriculum appear below. For example, a False Negative error was scored if a participant's measured BAC was .06 or higher, but the officer reported fewer than four signs. A False Positive error, for instance, was scored if the participant's BAC was below .06, but the officer reported six signs. A "Hit" occurred when the number of reported signs for a given BAC fell within the range reported below.

BAC	Number of Signs
≥ .06	4 - 6
.05 – .059	2 - 4
.03 – .049	0 - 4
≤ .03	0 – 2

Officers' observations of LSP as a function of stimulus speed were examined separately from total scores. At the two-second speed, LSP was reported for both eyes for all participants, (Tables 10 and 11). A breakdown of pursuit movements is not expected at very low BACs, and it is interesting that it was already observed at .016 and .019 with both the two-second and the one-second speeds. These apparent errors occurred at the fourth test period. By that time, the participants had undergone 8 examinations: 2 conditions x 4 test times. The data at hand do not provide a conclusive answer.

Table 11
Errors (Number) by Stimulus Speed and BAC

	2 sec (standard)		1 sec	
BAC (g/dL)	False Negative	False Positive	False Negative	False Positive
.097			1	
.088			1	
.077			1	
.076			1	
.075	1		1	
.063			1	
.063			1	
.063	1		1	
.059			1	
.019				1
.016		1		1
Total Errors	2	1	9	2
Total Errors by Condition	3		11	
Total Number of Scores	36		36	

The cases in which (1) total scores and LSP scores differed between conditions (i.e., between officers), and cases in which (2) total scores and LSP scores were incongruent with BACs were identified and examined case by case. In 13 examinations out of 72 total examinations, observations of LSP differed by condition and officer (i.e., between 1- and 2-second speeds). In all cases, LSP was observed in both eyes in the two-second condition, but there were no observations of LSP in the one-second condition.

The disruptions of the ability to track a moving stimulus and the observations of LSP at .063 to .097 BAC in the two-second condition are expected findings (Table 12). The parallel failures to observe LSP in the one-second condition at these BACs, therefore, are interpreted as false negative observations, and they can be attributed to the difficulty of viewing the eyes with such fast stimulus movement.

Table 12
Differences in Observations of LSP by Stimulus Speed and BAC

BAC (g/dL)	**Participant**	**Examination**	**Condition 1 (2 sec)**	**Condition 2 (1 sec)**
.097	5	1	2	0
.077	9	2	2	0
.073	4	4	2	0
.063	3	3	2	0
.063	8	2	2	0
.063	9	1	2	0
.059	5	3	2	0
.055	3	2	2	0
.048	9	4	2	0
.044	3	4	2	0
.042	5	4	2	0
.039	8	3	2	0
.030	2	4	2	0

For a more detailed examination of the potential source of the error in LSP, we reviewed the seven cases in which the BAC was less than .060. As can be seen in Table 12, LSP was reported only in the 2-second condition and not in the 1-second condition. The review of the videotapes revealed that, with one single exception, LSP could be observed in the videotaped examination. The exception was participant 8's observation at BAC = .039, which was not visible in the videotape.

In summary, false negatives (i.e., failures to detect LSP) were associated with variations in stimulus speed. The finding that rapid stimulus movement – with total movement duration of only one second - lessens the likelihood of observing LSP is relevant to law enforcement. In the interest of accurate roadside assessments, stimulus speed should not be faster than two seconds. However, because the errors are predominantly false negatives, and not false alarm errors, their implication is less critical for enforcement.

2. The Effects of Stimulus Elevation

During SFST training, officers are instructed to elevate the stimulus slightly above eye level to ensure that the eyelids are opened wide enough to allow a clear view of the eyes. Three specific elevations of the stimulus were examined to determine whether variations exerted negative effects on *occurrence* or *observation* of HGN signs. The data collected in this study used the same methodology that was used in the study on variations of stimulus speed.

The standard condition was defined for the purposes of the experiment as a stimulus position two inches above eye level. In the experimental variations, the stimulus was held below the standard elevation, at eye level, and above the standard elevation at four inches above eye level. Three officers were assigned by random procedure to the three conditions, and each demonstrated to the investigator's satisfaction that he could produce the assigned elevation. The obtained data are summarized in Table 13.

In the statistical analysis of equivalence, non-equivalence of the stimulus elevations was found when the observations were summed over time and decreasing BACs. More signs were observed when the stimulus was held at eye level or elevated four inches than when elevated two inches. However, in an analysis of data from the first test period only, no significant differences were found between the three elevations, and it can be noted that no errors occurred under any condition during the first test period.

Table 13
HGN Signs (Total Number) by Variations of Stimulus Elevation,
BACs, and Examination Period

			Stimulus Elevation (relative to eye level)		
BACs (g/dL)	**Exam**	**Participant's BAC (g/dL)**	**2" (standard)**	**0"**	**4"**
≥ .100	**1**	.109	6	6	6
		.102	6	4	6
		.100	6	6	4
		.100	6	6	6
.05-.099	**1**	.099	6	6	6
		.092	6	6	6
		.069	4	4	4
		.068	6	4	4
		.060	4	6	6
	2	.094	6	6	6
		.089	6	6	4
		.086	4	6	6
		.081	6	6	6
		.080	4	4	4
		.074	4	4	4
		.069	2*	2*	4
		.059	4	6**	4
	3	.091	2*	0*	6
		.074	4	4	4
		.065	0*	0*	4
		.063	4	4	2*
		.063	4	6	6
		.060	0*	4	4
		.059	0*	0*	6**
	4	.072	2*	4	4
		.052	6**	2	4
<.05	**2**	.047	6**	6**	6**
	3	.047	0*	4	0*
		.031	4	4	4
	4	.019	4**	4**	6**
		.033	0	4	4
		.034	0	0	4
		.038	0	4	4
		.038	0	4	4
		.047	0	4	4
		.049	4	4	4

*False Negative (FN) and **False Positive (FP) relative to the ranges specified above for the various BACs

A question of interest is whether the observation of more signs during test periods two, three, and four can be interpreted as higher accuracy due to stimulus positioning at eye level or four inches above eye level. Are the differences due to false negatives or false positives?

[11] Evidential instruments and the AlcoholSensor IV were used depending on the instruments available at each site.

Specifically, were signs incorrectly reported at BACs that are not expected to produce HGN? Or were signs missed? To address this issue, errors were examined case by case.

The number of false positive errors differed by time period and BAC (which were correlated), but they did not differ by stimulus position. The nine false positives were equally distributed across conditions; i.e., three at each elevation (Table 14). None occurred at the highest BACs during the first examinations. Four occurred in the second test period, one in the third test period, and four in the fourth test period. A videotape review revealed only one disagreement with the officers' scores. Their scores of four or six points for a participant with a .019 BAC at the fourth examination could not be confirmed by looking at the videotapes.

False-negative errors occurred most frequently with the standard elevation of the stimulus and during the third test period. Seven occurred with the stimulus elevated two inches, four with it at eye level, and two with a four-inch elevation. Two false negative errors occurred during the second test period, ten during the third period, and one during the fourth period. Thirteen of the officers' scores are judged to have been false negatives (i.e., they did not report the signs expected at the participants' BACs). The four-inch stimulus position opens the participant's eyes widely, a position that would be expected to minimize failures to detect HGN signs. There is no clear explanation for more errors at the two-inch position than at the eye-level position, and it is assumed that the finding reflects between-examiner differences in skill or criterion for reporting a sign.

Table 14
Errors (Number) by Stimulus Elevation and BAC

	2-inch elevation		Eye level		4-inch elevation	
BAC (g/dL)	False Negative	False Positive	False Negative	False Positive	False Negative	False Positive
.091	1		1			
.072	1					
.069	1		1			
.065	1		1			
.063					1	
.060	1					
.059	1		1	1		1
.052		1				
.047		1		1	1	1
.038	1					
.019		1		1		1
Total Errors	7	3	4	3	2	3
Total Errors by Condition	10		7		5	
Total Number of Scores	36		36		36	

Nine scores are judged to have been false positive (i.e., the participants' alcohol levels are not expected to produce the signs reported). The four-inch stimulus elevation, which widely opens the eyes, did not increase false positive observations in comparison to the other conditions. The data provide no evidence that this position engages different eye muscles than more moderate positions and, therefore, yields radically different observations.

3. The Effects of Stimulus Position in Front of the Face

Officers are trained to hold the stimulus approximately 12" to 15" in front of the suspect's face, a focal distance that affords a comfortable viewing of the stimulus. It is important to note the use of the word *approximately*, which acknowledges that officers conducting an examination at roadside cannot readily measure the exact distance between the stimulus and the suspect's face. Although there have been no reports that moderate variations in the distance affect either the *occurrence* or the *observation* of HGN signs, the issue had not been examined previously under controlled conditions.

The effects of varying the distance between the stimulus and the participants' faces were examined in three sessions with nine alcohol-dosed participants. At each session the stimulus was held at a different position: 12" – 15" in front of the face (the standard position), at 10", and at 20". The data was collected using the same procedure as the study on variations in stimulus speed and on variations of stimulus elevation. The results are summarized in Table 15.

Table 15
HGN Signs (Total Number) by Stimulus Distance from Face, BAC, and Examination

			Stimulus Distance (in front of face)		
BACs (g/dL)	**Exam**	**Participant's BAC (g/dL)**	**12-15" (standard)**	**20"**	**10"**
≥ .100	**1**	.129	6	6	6
		.112	4	4	4
		.106	4	6	4
		.104	4	6	6
		.102	6	6	6
.05-.099	**1**	.091	6	6	6
		.087	4	4	4
		.079	6	6	6
		.070	6	4	6
		.066	2*	4	4
	2	.085	4	6	4
		.079	2*	2*	6
		.076	6	6	6
		.075	2*	4	6
		.064	6	4	6
		.059	4	2	4
		.052	4	4	4
	3	.070	4	4	6
		.063	2	4	4
		.057	2	4	4
		.051	4	4	6**
<.05	**1**	.044	4	4	4
	2	.045	6**	6**	6**
		.036	4	4	4
	3	.049	4	4	4
		.035	2	4	4
		.029	6**	4**	6**
		.025	4**	4**	2
	4	.046	2	2	4
		.046	4	4	6**
		.037	0	2	0
		.036	6**.	4	4
		.027	4**	4**	4**
		.022	4**	4**	4**
		.120	4**	4**	6**

*False Negative (FN) and **False Positive (FP) relative to the ranges specified above for the various BACs

Statistical analyses of equivalence were conducted for total scores across test periods and also for scores from the first test period only. Statistical equivalence was supported for the standard (12" – 15") versus the more distant stimulus position (20"), indicating that there was not a statistically significant difference in the number of HGN signs observed in those two conditions. However, for the standard position versus the 10" position, the null hypothesis of equivalence was rejected. Officers reported more signs at the reduced distance than at the standard distance. The question that must be answered then is whether the observations of more signs are correct observations or whether they are false positive errors. Table 16 provides a breakdown of the errors. As can be seen in this table, the most frequent errors were false positives, and their frequencies were essentially identical in the three different conditions. Thus, the greater number of observations at the 10” position cannot be attributed to false positives.

Table 16
Errors (Number) by Stimulus Distance from Face and BAC:

	12 – 15”		**10”**		**20”**	
BAC (g/dL)	False Neg.	False Pos.	False Neg.	False Pos.	False Neg.	False Pos.
.079	1		.		1	
.075	1					
.066	1					
.051						
.046				1		
.045		1		1		1
.036		1		1		
.029		1				1
.027		1		1		1
.025		1		1		1
.022		1		1		1
.012		1		1		1
Total Errors	3	7		7	1	6
Total Errors by Condition	10		7		7	
Total Number of Scores	36		36		36	

No false-negative errors occurred with the 10” distance, one occurred with the 20” distance, and three occurred at the standard 12” – 15” distance. The between-condition difference in false negatives indicates that the non-equivalence finding reflects somewhat more accurate viewing of eye signs at a closer distance. Note, however, that it may also indicate between-officer differences in skill or in the individual officer’s criterion for observations, and the small gain must be weighed against risks to the officer as a result of standing closer to a suspect.

The numbers of observational errors were closely similar for all three signs (LSP, MAX, and AOG) with no single sign accounting for a disproportionate number of errors.

III.1.C Summary: Experiment 1 Findings

In the controlled environment of a laboratory experiment, trained and skilled officers did not make errors during their observations of participants at BACs ≥ .10. At those alcohol levels, variations of stimulus speed, elevation, and distance from the participants' face made no measurable difference in the accuracy of their observations. When participants' mean BAC was .088, officers made only four false-negative errors – failing to detect impairment - during the first of four examination periods.

a. Stimulus speed
Data obtained in the laboratory demonstrate that rapid stimulus movement (a one-second pass) significantly decreases an officer's ability to detect the HGN sign. The optimal viewing time for the stimulus movement (center-to-side and side-to-center) was shown to be approximately two seconds. Slow movement (four seconds) neither increased observational errors nor improved observations.

b. Stimulus elevation
During the first test period when the mean BAC for nine participants was .088 (and ranged from .060 to .109), varied elevations of the stimulus, eye-level versus two or four inches above eye level, produced equivalent observations of HGN signs. Three additional examinations on a descending alcohol curve produced significant differences as a function of elevation. The officer who held the stimulus two inches above eye-level observed significantly fewer signs than the officer who held the stimulus at eye level. A four-inch elevation produced the smallest number of false negatives *without increasing false positives*, and the data provide no evidence that the position increases errors by engaging different eye muscles than those engaged at a more moderate position.

c. Stimulus distance from face
The standard for the distance between the stimulus and the Participant's face is approximately 12" - 15". Increasing that distance to 20" did not alter the number of signs observed. When the distance was decreased to 10", officers correctly reported more signs. The magnitude of the difference is small. However, the magnitude of the difference due to the increase in correctly reporting signs is small. Almost two-thirds of the false positive errors of the experiment occurred during the sessions when stimulus distance was varied. The errors, which were distributed equally across conditions, possibly reflect officer skill and/or participant characteristics. The finding should be viewed cautiously pending replication.

III.2 Experiment 2

The verbal instructions for HGN examinations include, “Stand with your feet together and your arms at your side.” Although most examinations are conducted in this standing position, on occasion it is not possible to do so, because the suspect cannot stand. A crash victim who has been placed on a gurney must be examined in a supine position. A driver who is confined to a wheelchair must be examined in a seated position. Since these circumstances preclude balance and walking tests, questions about possible effects of position on HGN signs are important both for the officer and the suspect. In this experiment, officers examined alcohol-dosed participants in standing, sitting, and lying-down positions. Two hypotheses were tested using statistical tests of equivalence (see Appendix III) and they are formalized in Table 17.

Table 17
Variable and Hypotheses of Experiment 2

Variable	Hypothesis
Participant Position	The *occurrence* of and SFST-trained officers’ *observations* of HGN signs when a Participant is seated = The *occurrence* and SFST-trained officers’ *observations* of HGN signs when a Participant is standing
	The *occurrence* and SFST-trained officers’ *observations* of HGN signs when a Participant is lying down = The *occurrence* and SFST-trained officers’ *observations* of HGN signs when a Participant is standing

III.2.A Sites for Data Collection

In response to a notice on a computer-based network of certified DREs (the DRE listserv), training agency personnel in four western States volunteered to participate in the research project. They agreed to permit data collection during the alcohol workshops that are an integral part of SFST and DRE training. SCRI staff traveled to workshops held in Arizona, Texas, and California. Data collection in Medford, Oregon, was done under the supervision of a senior Oregon trooper. The total number of workshops conducted at each site is presented in Table 18.

Table 18
Alcohol Workshop Sites

Location	Number of Workshops
Medford, Oregon	2
Phoenix, Arizona	1
Prescott, Arizona	1
Sacramento, California	2
San Antonio, Texas	2

III.2.B Workshop Participants and Procedures

The collection of data in the workshop setting was contingent on the willingness of the SFST instructors to assist with the research. At all sites where data were obtained, SFST-qualified officers volunteered as examiners for this study.

At the first site in Prescott, Arizona, data were obtained following the initial breath testing[11] of drinking participants and again toward the end of the workshop. That experience demonstrated that two periods of access to the participants took an excessive amount of time away from training activities. Although the instructors were fully cooperative, the investigator was reluctant to continue data collection in a manner that disrupted training. Therefore, at the Phoenix, San Antonio, and Sacramento workshops HGN observations were obtained only once at the end of the drinking-absorption period. This manageable timing allowed the collection of complete data sets with minimal disruption to training.

As the first activity of data collection, baseline data (i.e., HGN examinations with participants in the standing position) were obtained by two officers who *did not* participate in data collection in the sitting and lying-down position. After the standing-position examinations were completed, half of the participants were instructed to sit and half were instructed to assume a supine position. (Cots or mats were available at two sites. In other locations, participants reclined on desks or tables.) The remaining officers were summoned to test all participants. The participants were instructed to change positions for the second round of examinations. Officers recorded their scores with the form that can be seen in Appendix IV. As the forms were completed, they were handed to the SCRI personnel on site for immediate review. This allowed time for the few omissions that were noted to be brought to the attention of the officer for correction.

III.2.C Results: Experiment 2

Eight workshops, 77 officer-examiners, and 75 drinking participants produced a total of 907 completed data forms. Appendix V lists all instructors, officers, and other personnel who participated in the workshops. Participant characteristics are detailed in Appendix VI, and the number of officers, workshops, participants and completed data forms at each site are summarized below in Table 19.

Table 19
Summary: Field Data Collection

Location	Examiners (No.)	Workshops (No.)	Participants (No.)	Completed Data Forms (No.)
Medford, OR	12	2	18	179
Phoenix, AZ	13	1	10	88
Prescott, AZ	28	1	18	144
Sacramento, CA	12	2	16	367
San Antonio, TX	12	2	13	129
Total	77	8	75	907

The results from each location are presented in Table 20, and the results of the pooled data are provided in Table 21. Figure 2 illustrates the officers' performance in terms of percentage of correct responses, false positives, and false negatives, for each of the three participants' positions. By site, 80% to 97% of HGN observations were correct (i.e., the observations reported by the officers were consistent with participants' measured BACs). The seated position proved to be the most difficult for the examinations and yielded the fewest (88%) correct HGN scores. However, whether the participant was standing or lying down had little differential effect on the officers' reports of HGN signs. Ninety-two percent of observations were correct when the participants stood. Ninety-one percent were correct when they were lying down. The few errors that the officers made were mostly due to failing to report HGN signs (7% false negatives) than to reporting signs not supported by BACs (3.4%).

Analysis revealed that more than 40% of the variability in HGN observations was accounted for by significant differences between-participants, and a best-fit model of the effects of *standing versus sitting* and *standing versus lying down* included individual differences as a random effect and BAC as a covariate. Using this model, the upper and lower bound non-equivalence tests led to rejection of the null hypotheses. Thus, a claim of equivalence of the three positions is supported. Details of the equivalence analysis are given in Appendix VII.

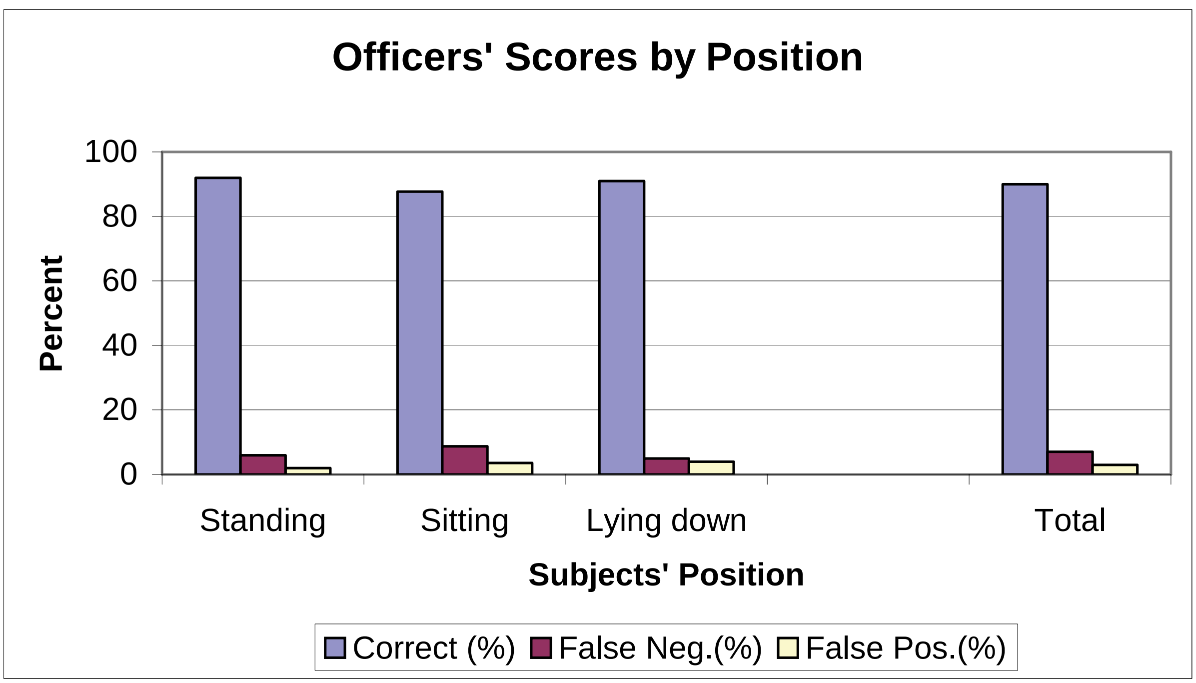
Table 20
Summary of HGN Observations by Site

BAC (mean)	Position	Correct	False Negatives	False Positives	Total Exams
Medford, OR					
.107	Standing	33	1	1	35
	Sitting	82	6	0	88
	Lying	47	2	7	56
	Totals	162	9	8	179
	% of Total Exams	90.5%	5.0%	4.5%	
Prescott, AZ					
.091	Standing	19	1	0	20
	Sitting	57	5	13	75
	Lying	44	2	3	49
	Totals	120	8	16	144
	% of Total Exams	83.3%	5.6%	11.1%	
Phoenix, AZ					
.093	Standing	32	4	0	36
	Sitting	23	8	0	31
	Lying	15	6	0	21
	Totals	70	18	0	88
	% of Total Exams	79.5%	20.5%	0	
Sacramento, CA					
	Standing	29	2	0	31
	Sitting	162	6	0	168
	Lying	165	3	0	168
	Totals	356	11	0	367
	% of Total Exams	96.7%	3.3%	0	
San Antonio, TX					
.	Standing	23	1	2	26
	Sitting	38	11	2	51
	Lying	44	5	3	52
	Totals	105	17	7	129
	% of Total Exams	81.4%	13.2%	5.4%	

Table 21
Summary: HGN Observations by Participants' Positions

Position	Correct	False Negatives	False Positives	Total Exams
Standing	136	9	3	148
	91.9%	6.1%	2.0%	
Sitting	362	36	15	413
	87.7%	8.7%	3.6%	
Lying	315	18	13	346
	91%	5.2%	3.8%	
Totals	813	63	31	907
	89.6%	6.9%	3.4%	

Figure 2
Officer Performance by Participant Position (Standing, Seated, Lying Down)



III.2.D Summary: Experiment 2 Findings

The results from an analysis of more than 900 examinations provide no compelling evidence that a suspect's position while being examined for HGN is a critical variable. These findings are consistent with the study conducted by Citek, Ball, and Rutledge (2003) which found that standing, seated, and supine positions had no effect on the detection of impairment through HGN at BACs of .08 and .10. Citek et al. used a similar procedure where participants from law enforcement alcohol workshops were utilized to test the effect of position on HGN signs at varying levels of alcohol impairment.

Although both the data and the officers' comments in the current study suggest that having participants sit is somewhat less favorable than having them stand or lie down, the difference is neither large nor statistically different. The obtained data support the validity of HGN observations whether the participant is standing, sitting, or lying down and provide no reason to change current practices.

III.3. Experiment 3

Officers are trained to examine both of a suspect's eyes for signs of HGN and to score the three possible signs separately for each eye. On occasion, however, an individual may inform an officer that he or she has no vision or very limited vision in one eye. The condition can be the result of prosthesis, amblyopia, severe presbyopia, or other medical condition. Whatever the etiology, the condition raises questions concerning the validity of HGN signs. Specifically, are HGN observations valid if the person has the use of only one eye?

During a preliminary discussion of the issue, Dr. Karl Citek, Pacific College of Optometry, expressed the view that HGN in a functioning eye is not affected by a non-functioning eye. Also, the following comments were posted on a computer-based network of certified DREs (the DRE listserv) by Dr. Morris Odell, Forensic Physician, Victorian Institute of Forensic Medicine, Australia. Since Dr. Odell is discussing unilateral nystagmus (i.e., the occurrence of nystagmus in only one eye without reference to whether one or both eyes are functional), his comments are tangential to the issues of the experiment, but of interest nonetheless. Dr. Morris commented that:

> Unilateral nystagmus is never normal. This could indicate some medical condition involving the vestibular/oculomotor pathways, or a cerebellar problem. There is a multitude of possible causes including some very nasty ones, as well as benign conditions. I certainly wouldn't be using it as evidence of alcohol or drug intoxication. I think anyone with an abnormal sign like that should be sent for medical evaluation.

Table 22 presents the hypothesis of Experiment 3.

Table 22
Hypothesis of Experiment 3

Variable	Hypothesis
Functional Vision	HGN signs observable in the eyes of an individual with functional vision in one eye only = HGN signs associated with a specific BAC.

III.3.A Participants

It proved moderately difficult to locate individuals who met all of the experiment criteria. In addition to the standard requirements of SCRI experiments concerning health status, drug/alcohol use, and availability, applicants were required to be functionally one-eyed, willing to consume a moderate amount of alcohol, and willing to participate in a single six- to eight-hour weekday session. Recruitment efforts through contacts with medical facilities and a college of optometry yielded no candidates. A number of people, however, responded to a posting on an Internet listserv, and telephone interviews identified eight individuals who appeared to meet the study criteria. They were scheduled for a single alcohol session. Due to the limited number of participants, this study is a ***preliminary*** analysis on the effects of non-functional eyes on HGN examinations.

Because participants to be given alcohol in SCRI experiments are transported from their residences to the laboratory by taxi, and because fares from distant locations in the city could be prohibitive, participants were required to live within a 10-mile radius of the facility. The individuals who qualified for Experiment 3 all lived outside that area making it necessary to hold the session offsite at a location that minimized the travel distance.

Participant 7, a female, did not appear for the scheduled session. Efforts to contact her failed, because she had provided an incorrect telephone number. Therefore, data were obtained from only seven individuals.

Acting on clinical guidance from Dr. Ronald Matsumoto, Los Angeles College of Optometry, the vision criterion for participants was set as "*corrected* vision in one eye of less than 20/100". The underlying rationale for the criterion, as explained by Dr. Matsumoto, is that *uncorrected* vision of less than 20/100 indicates poor vision, but if corrective lenses improve the individual's vision above that level, he or she is not considered functionally one-eyed. Therefore, the standard as given above is stated in terms of *corrected* vision.

To confirm that participants' vision qualified them to participate in the experiment, they were vision-tested upon arrival for the session using a standard Snellen's chart. The chart's block letters' size decreases line-by-line. At the appropriate distance, each letter in the 20/20 line subtends a visual angle of 5 degrees and each component part subtends an angle of 1 minute.

The vision of Participants 1, 3, 5, and 8 was measured as less than 20/200 (i.e., from a distance of 20 feet, they could not read the top line of the Snellen chart). This level of vision (20/200) is also the threshold for legal blindness. Each of these individuals spontaneously stated that he was "blind" in the affected eye. Participant 4's vision in the left eye was less than 20/100. Participant 2 has an artificial right eye as a consequence of injury to his eye at age two. Participant 6's *uncorrected* vision was measured as less than 20/100 in both eyes, but he had *corrected* vision of 20/50 in both eyes and thus did not strictly meet the vision criterion. Given the small number of qualified participants at the session, he was retained as a participant and data were obtained without his corrective lenses.

III.3.B Alcohol Treatments

Six participants were given alcohol (80-proof vodka and orange juice) with doses calculated on the basis of body weight to produce peak BACs of .10. Participant 8, a light drinker, was given alcohol to produce a .08 BAC. Although he claimed vodka as his preferred beverage, he was unable to tolerate the dose, experienced nausea, and reached a peak BAC of .062.

Alcohol beverages, given as three equal drinks at 10-minute intervals, were consumed over a 30-minute period. Following a 30-minute absorption period, BAC measurements were obtained with an AlcoSensor IV breath test device.

III.3.C HGN Examinations

Two certified DREs independently examined participants and recorded their observations of HGN signs. It was obvious to any observer, including the officer-examiners, that participant 2's right eye was artificial. For the other participants, the officers had no knowledge about the participants' eyes (i.e., whether one or which one was non-functioning).

Immediately following breath testing, the first participant was escorted to a private room for the HGN examination by Officer 1. When Officer 1 completed his observations, the participant was escorted to the second room where Officer 2 conducted his examination. The participant then was escorted back to the lounge area. A second breath specimen was obtained by AlcoSensor, and the participant was given lunch. This order of events proceeded for all participants. Exactly one hour after each participant's first examination, BAC measurements and HGN observations were repeated for Examination 2. Participants remained in the facility until their BACs decreased below .03, and then were transported to their residences by taxi.

III.3.D Results: Experiment 3

Table 23 details the results of each test for each participant. The mean scores for each participant are summarized in Table 24. Because participant 2's artificial eye could not display HGN signs, his scores were not included in the statistical analysis. The officers observed the same signs in both eyes for 19 of the remaining 24 examinations. Officer 2 observed fewer signs in the non-functioning eye for 5 of the 16 examinations.

Table 23
HGN Scores for One-Eyed Participants (N=7)

			Examination 1											
			Officer 1						Officer 2					
			Left Eye			Right Eye			Left Eye			Right Eye		
S	Eye[1]	BAC(g/dL)	LSP	MAX	AOG	LSP	MAX	AOG	LSP	MAX	AOG	LSP	MAX	AOG
1	R	.091	1	1	1	1	1	1	1	1	1	1	1	-
2	R[2]	.094	1	1	1	-	-	-	1	1	1	-	-	-
3	L	.117	1	1	1	1	1	1	-	-	-	-	-	-
4	L	.085	1	1	1	1	1	1	1	-	-	1	-	-
5	L	.071	1	-	-	1	-	-	1	1	1	1	1	1
6	R	.104	1	1	1	1	1	1	1	1	1	1	1	1
8	L	.062	1	1	1	1	1	1	1	1	-	1	1	1
		Mean	1	0.9	0.9	0.9	0.7	0.7	0.9	0.7	0.6	0.7	0.6	0.4
		α	0	0.4	0.4	0.4	0.5	0.5	0.4	0.5	0.5	0.5	0.5	0.5
			Examination 2											
1	R	.064	1	1	1	1	1	1	1	1	1	1	1	-
2	R[2]	.072	1	1	1	-	-	-	1	1	1	-	-	-
3	L	.083	1	-	-	1	-	-	-	-	-	-	-	-
4	L	.079	1	1	-	1	1	-	1	-	-	1	1	1
5	L	.056	1	1	1	1	1	1	1	1	1	1	1	1
6	R	.068	1	1	1	1	1	1	-	1	-	-	1	-
8	L	.060	1	1	-	1	1	-	1	1	-	1	1	1
		Mean	1	0.9	0.6	0.9	0.7	0.4	0.7	0.7	0.4	0.6	0.7	0.4
		α	0	0.4	0.5	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

1 Eye that met "non-functioning" criterion. Measured vision was 20/100 or poorer.
2 Artificial right eye

Table 24
Total Nystagmus Scores (Mean of Two Officers)
By Participants' BAC and Functional Eye Status

Examination 1			
Participant	BAC (g/dL)	Functioning Eye	Non-Functioning Eye
1	.091	3	2.5
2	.094	3	0
3	.117	1.5	1.5
4	.085	2	2
5	.071	2	2
6	.104	3	3
8	.062	3	2.5
	Mean	2.5	1.9
	α	0.6	1.0
Examination 2			
Participant	BAC (g/dL)	Functioning Eye	Non-Functioning Eye
1	.064	3	2.5
2	.072	3	0
3	.083	0.5	0.5
4	.079	2.5	1.5
5	.056	3	3
6	.068	2	2
8	.060	2	2.5
	Mean	2.3	1.7
	α	0.9	1.1

The mean number of HGN signs for functioning eyes was 2.33. For non-functioning eyes, it was 2.08. Although the difference is small, it is statistically significant (paired $\underline{t}$ 2.30, 23 df, $p<0.03$), and due to the observation of fewer signs in the non-functioning eye.

With one exception, the mean scores for the functioning eyes are within the expected range. The one exception was participant 3, for whom the officers reported fewer than expected signs at the .117 BAC. However, one of the officers reported three signs in each eye for this Participant at the first test period, and therefore this one case provides no basis for concluding that nystagmus in a functioning eye is affected by a non-functioning eye.

III.3.A Summary: Experiment 3 Findings

The data obtained with a sample of seven participants yielded a small but statistically significant difference between functioning and non-functioning eyes that is attributable to *fewer* HGN signs in non-functioning eyes. While the data is preliminary, it suggests that HGN examinations of individuals with monocular vision should not yield misleading information.

Again, the findings should be viewed as preliminary because it is unknown whether they extend to individuals with different underlying medical causes for non-functioning eyes (Criden & Ellis, 2007; Holt & Hold, 1988, Wachler & Holds, 1998; Yee, Cravens & Kotler, 1975), and because the sample size was small. Additional research needs to be conducted to verify this finding.

IV. GENERAL DISCUSSION

Laboratory experiments and data collection in field settings examined the effects of variations in the administration of HGN. Forty participants were examined in laboratory experiments and 75 participants were examined during alcohol training workshops.

Variations of the stimulus movement and of the participants' positions were evaluated using statistical tests of equivalence, and were found to have minimal effects on the officers' *observations* and no measurable effects on the actual *occurrence* of HGN signs. The few discrepancies that were noted between measured BACs and observed signs are not sufficient to support the conclusion that HGN signs are altered by the tested variations in administration.

Data obtained from one-eyed individuals revealed a small but statistically significant difference between functioning and non-functioning eyes. The difference is attributable to one officer's report of fewer signs in the non-functioning eyes of some of the participants. It is not possible with the available data to determine whether a between-officer difference in criteria for *observations* accounts for the finding and/or whether HGN signs *occurred* less distinctly in the non-functioning eyes.

It is not possible to broadly generalize the findings from the small sample of individuals with only one functioning eye. Given that different conditions cause eyes not to function, questions about HGN in functionally monocular drivers could be fully resolved only with a large experiment. In view of the following considerations, such effort may not be merited.

- It is estimated that 3.1% of Americans of all ages have blindness in one or both eyes (Leonard, 2002). Therefore, the likelihood of someone with monocular vision who comes to the attention of traffic officers for suspicion of DUI is relatively small.
- Officers routinely record suspects' responses to pre-test queries about medical problems. Assuming that suspects acknowledge their visual limitations, that information then becomes part of the total information underlying the interpretation of the observations.
- Because HGN appears to be reduced in a non-functioning eye, if officers were to rely solely on eye signs, they would only increase their false-negative rates and they might improperly release one-eyed individuals. There is no evidence that HGN signs in such individuals will lead to false arrests.

V. CONCLUSIONS

A one-second stimulus speed significantly increased the number of false-negative errors.

In the laboratory experiments, the officer-examiners did not err at BACs $\geq$ .10 and rarely erred at BACs $\geq$.08.

> *SFST-trained and experienced officers did not err in their reports of HGN signs when participants BACs were .10 or higher, and they made very few errors at $\geq$.08 BAC.*

Officers correctly observed more HGN signs when the stimulus was held 10” in front of participants’ faces than when it was positioned 12” to 15” or 20” from their faces.

> *Although there were fewer failures to observe HGN signs with a close viewing distance (10”), the magnitude of the difference was not large, and the gain of fewer errors – and reduction in false negatives – must be weighed against a possible officer safety issue that might result from standing so near to suspects.*

Participants’ positions (standing, sitting, or lying down) had no statistically significant effects on officers’ reports of HGN signs.

In conclusion, HGN as used by law enforcement is a robust procedure. The study findings provide no basis for concluding that the validity of HGN is compromised by minor procedural variations.

References

Bobo, J. (2003). Foundational hurdles of HGN. *Admissability of Horizontal Gaze Nystagmus Evidence: Targeting Hardcore Impaired Drivers.* Alexandria, VA: American Prosecutors Research Institute.

Burns, M., & Moskowitz, H. (1977). Psychophysical Tests for DWI Arrest. (Publication No. DOT-HS-802-424). Washington, DC: National Highway Traffic Safety Administration.

Cahalan, D., Cisin, I. H., & Crossley, H. M. (1969). *American Drinking Practices: A National Study of Drinking Behavior and Attitudes*. New Brunswick: Rutgers Center of Alcohol Studies.

Citek, K., Ball, B., & Rutledge, D. A. (2003). Nystagmus testing in intoxicated individuals. *Optometry*, *74*, 695-710.

Criden, M. R., & Ellis, F. J. (2007). Linear nondisplaced orbital fractures with muscle entrapment. *Journal of American Association for the Pediatric Opthalmology and Strabisumus, 11*, 142-147.

Holt, G. R., & Holt, J. E. (1988). Management of orbital trauma and foreign bodies. *Otolaryngologic Clinics of North America, 21*, 35-52.

Leondard, R. (2002). Statistics on Vision Impairment: A Resource Manual. Retrieved from http://www.gesta.org/estudos/statistics0402.pdf

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Pharmacodynamics, 15*, 657-680.

Stuster, J., & Burns, M. (1998). Validation of the Standardized Field Sobriety Test Battery at BACs Below .10 Percent. (Publication No. DOT HS 808 839). Washington, DC: National Highway Traffic Safety Administration.

Tharp, V., Burns, M., & Moskowitz, H. (1981). Development and Field Test of Psychophysical Tests for DWI Arrest. (Publication No. DOT HS 805 864). Washington, DC: National Highway Traffic Safety Administration.

Wachler, B. S., & Holds, J. B. (1998). The missing muscle syndrome in blowout fractures: an indication for urgent surgery. Ophthalmic Plastic and Reconstructive Surgery, 14, 17-8.

APPENDICES I - VII

Appendix I

Laboratory Experiment Participants' Characteristics

Participant No.	Gender	Age	Height (in)	Weight (lbs)
1	Female	24	61	105
2	Female	29	63	141
3	Male	26	69	170
4	Male	29	74	220
5	Male	29	69	160
6	Male	27	69	155
7	Female	27	65	145
8	Male	43	76	195
9	Female	32	69	130
10	Male	35	70	165
11	Female	47	65	148
12	Male	23	75	185
13	Female	37	65	155
14	Male	28	69	155
15	Male	47	73	200
16	Female	27	62	105
17	Female	40	64	107
18	Male	30	75	200
19	Female	34	64	103
20	Female	25	69	140
21	Male	28	72	185
22	Male	40	69	260
23	Female	43	69	138
24	Male	54	71	169
25	Female	26	65	150
26	Male	26	69	154
27	Male	25	70	195

Appendix II

Officer-Examiners for the Laboratory Experiment

Officer	Agency	Training/Qualifications
Sergeant Rick Colbert	Long Beach Police Dept.	
Sergeant Gordon Collier	Long Beach Police Dept	
Officer Eric Frank	Glendale Police Dept.	
Staff Sgt. Gary Greenbush	U.S. Marine Corps	
Officer Clark John	Los Angeles Police Dept.	
Officer Rich Ulrich	Glendale Police Dept.	
Sergeant Helena Williams	California Highway Patrol	

Appendix III

REPORT OF THE ANALYSIS OF LABORATORY EXPERIMENT DATA

Barbara Tabachnick, Ph.D.

I. Analytic Approach

For all analyses, the six indicators of officer-observed HGN (three signs for each eye) yielded a scale varying from zero (no indicators) to six (all three signs in both eyes). These HGN scores were recorded for each of four exams in all conditions. Thus, for each of the 3 variables that were examined, 9 participants provided a maximum of 36 HGN scores for each combination of conditions (e.g., standard versus increased speed). Distributions of observed HGN scores were acceptable given the sample sizes for all analyses.

Analyses of covariance through SPSS provided information about HGN means and standard errors adjusted for BAC. These values were then used in the SOLAS software, which develops confidence intervals for differences between means as well as tests of equivalences of pairs of means.

Equivalence between means for a pair of conditions were evaluated in EquivTest (Statistical Solutions) using a Schuirmann OST/TOST procedure in which one-tailed *nonequivalence* tests are applied to each of two null hypotheses: one for an upper bound of nonequivalence between means and another for a lower bound. A meaningful difference using a criterion of ± 20% (a conservative industry standard) of the HGN scores in the baseline condition was used for the Schuirmann tests. Rejection of the null hypothesis for these tests in both directions indicates that equivalence can be claimed. ***Non-rejection of either test is associated with nonequivalence.***

Planned analyses included five comparisons: (1) increased versus standard speed, (2) reduced versus standard distance, (3) increased versus standard distance, (4) reduced versus standard elevation, and (5) increased versus standard elevation. All planned analyses were preceded by a “2 x 4” within-participants analysis of covariance (ANCOVA): two conditions by four exams with BAC as a time-varying covariate. None of the interactions were statistically significant ($p > .05$), so comparisons between conditions were averaged over the four exams for all planned analyses.

II. Variations in Rate of Speed of Stimulus Movement
Standard (2 seconds) versus increased (1 second) speed

One participant failed to yield a score in the fourth exam with increased speed, so 35 scores were available.

Stimulus speed effects were examined in a one-way within-participants ANCOVA with two conditions (standard versus increased speed) with BAC as a time-varying covariate. The 90% confidence interval for the differences in adjusted HGN between the two conditions was 0.981 to 1.905. The nonequivalence test of the lower 20% equivalence bound yielded t = -

2.24, *p* = .984, so that equivalence to within the specified bound cannot be claimed. A larger number of HGN signs was observed by officers in the standard condition (M = 4.17, SE = 0.192) than in the 1-second condition (M = 2.73, SE = 0.195). Post-hoc analysis evaluated data from the first exam (in which BAC was highest) in a one-way ANOVA between speed conditions. The 90% confidence interval for the differences in adjusted means between the two conditions was -0.108 to 2.998. The nonequivalence test of the lower 20% equivalence bound yielded *t* = -0.61, *p* = .721, so that equivalence to within the specified bound still cannot be claimed.

III. Variations in Elevation of Stimulus
Standard (2") versus reduced (0") elevation

Position effects were examined in a one-way ANCOVA with two conditions (standard versus reduced elevation) with BAC as a time-varying covariate. The 90% confidence interval for the differences in adjusted HGN between the two conditions was 0.113 to 1.323. For the nonequivalence test of the lower 20% equivalence bound, *t* = -0.065, *p* = .527, so that equivalence to within the specified bound cannot be claimed. More HGN signs were observed at the reduced elevation (M = 4.192, SE = 0.253) than at the standard elevation (M = 3.474, SE = 0.253).

A post-hoc analysis evaluated the first exam (in which BAC was highest) in a one-way ANOVA between positions. The identical means of 5.444 produced a 90% confidence of -0.368 to 0.368. A claim of equivalence is obviously supported with no difference between means.

Standard (2") versus increased (4") elevation

Position effects were examined in a one-way ANCOVA with two conditions (standard versus increased elevation) with BAC as a time-varying covariate. The 90% confidence interval for the differences in adjusted mean HGN between the two conditions was 0.540 to 1.682. For the nonequivalence test of the upper 20% equivalence bound, *t* = -1.22, *p* = .884, so that equivalence to within the specified bound cannot be claimed. More HGN signs were observed at the increased elevation (M = 4.611, SE = 0.239) than at the standard elevation (M = 3.500, SE = 0.239).

A post-hoc analysis evaluated the first exam (in which BAC was highest) in a one-way ANOVA between positions. The 90% confidence interval for the differences in HGN between the two conditions was -0.967 to 0.521. Tests of both upper and lower bounds supported rejection of the null hypotheses of nonequivalence, so that equivalence between conditions may be claimed.

IV. Variations in Distance of Stimulus from Participant's Face
Standard (12" - 15") versus reduced (10") distance

One participant failed to yield a score in the fourth exam with reduced distance and another failed to yield any scores in the fourth exam, so that 35 scores were available for standard and 34 for reduced distance conditions.

Distance effects were examined in a one-way ANCOVA with two conditions (standard versus reduced distance) with BAC as a time-varying covariate. The 90% confidence interval for the differences in adjusted HGN between the two conditions was 0.427 to 1.205. The nonequivalence test of the upper 20% equivalence bound produced $t = -0.016$, $p = .506$, so that equivalence to within the specified bound cannot be claimed. Officers reported observations of more HGN signs at the reduced distance (M = 4.88, SE = 0.165) than in the standard condition (M = 4.06, SE = 0.163).

A post-hoc analysis evaluated the first exam (in which BAC was highest) in a one-way ANOVA between distance conditions. The 90% confidence interval for the differences in HGN means between the two conditions was -0.103 to 0.991. For the nonequivalence test of the upper 20% equivalence bound, $t = -1.82$, $p = .053$, so that equivalence to within the specified bound still cannot be claimed to within the 20% criterion.

Standard (12" to 15") versus increased (20") distance

One participant failed to yield any scores in the fourth exam, so that 35 scores were available for each distance condition.

Distance effects were examined in a one-way ANCOVA with two conditions (standard versus increased distance) with BAC as a time-varying covariate. The 90% confidence interval for the differences in adjusted HGN between the two conditions was -0.353 to 0.469. Both the upper and lower bound nonequivalence tests yielded rejection of the null hypotheses, so that a claim of equivalence is supported.

V. Summary

Equivalence Evaluations

- When scores from four examinations were combined over diminishing levels of BAC, only one comparison (standard distance from the face versus increased distance) supported equivalence.

- With scores from the first examination only (i.e., at the highest mean BACs), the numbers of observed HGN signs were statistically equivalent over changes in stimulus elevation from 0" to 2" and from 2" to 4".

- Increased speed of stimulus movement produced a decrease in observed HGN signs at the first exam as well as in the combination of all four exams.

- Increased or decreased elevation of the stimulus, relative to eye level, produced an increase in observed HGN signs in the combination of all four exams.

- Decreased distance of the stimulus from the participant's face increased the number of observed HGN signs at the first exam as well as in the combination of all four exams.

Appendix IV

Data Collection Form for Alcohol Workshops

Date ____/_____/ 2003 Officer ____________________
Clock time __________
Participant ____________
Position (check one)
Standing_____
Sitting_______
Lying_______

HGN	Left Eye	Right Eye
Lack of Smooth Pursuit		
Max. Deviation		
Onset prior to 45 degrees		
Total		

Complete form only to this line.

Male _______ Female______ BAC _________

Age (est) 21 – 30_____
31 – 40_____
41 – 50_____
51 – 60_____
61+_____

Appendix V

Officer Rosters, Field Data Collection

CALIFORNIA HIGHWAY PATROL
SACRAMENTO, CA
JULY 29-30, 2003

Lieutenant Deborah Schroder Sergeant Helena Williams
Sgt. Robert Brunell
Officer Jason Craven
Officer Stacy Doyle-Barr
Officer Kevin Dwyer
Officer Scott Fredrick
Officer Vaughn Gates
Officer Jeff George
Officer Robert Hays
Officer Daniel Lamm
Sergeant Bill Languemi
Officer Kelly Lassey
Officer Gilbert Lee

OREGON STATE POLICE
MEDFORD, OR
MAY 18-20, 2003

Officer	Agency
Senior Trooper Ken Snook	Oregon State Police, Grants Pass
Officer Donovan Schmidt	Grants Pass Dept. of Public Safety
Officer Dennis Ward	Grants Pass Dept. of Public Safety
Sandy Nelson	Jackson County Sheriff's Office
Deputy Ralph Nelson	Jackson County Sheriff's Office
Officer Nathan Sickler	Klamath Falls Police Department
Officer Greg Lemhouse	Medford Police Department
Officer Jim Swanson	Medford Police Department
Officer Tim Lenihan	Myrtle Creek Police Department
Trooper Scott Holsworth	Oregon State Police, Grants Pass
Trooper Aaron Olympius	Oregon State Police, Grants Pass
Senior Trooper Tim Plummer	Oregon State Police, Roseburg
Officer Shane Wilson	Talent Police Department

NORTHERN ARIZONA REGIONAL TRAINING ACADEMY
PRESCOTT, AZ
MAY 2, 2003

Officer	Agency
Sergeant Dick Studdard	Los Angeles Police Department (ret.)
Officer Michael Boucher	Arizona DPS
Officer Steve Costello	Arizona DPS
Officer Kevin P. Jones	Arizona DPS
Officer Joshua Gonzalez	Arizona DPS
Deputy Oscar Berrelez	Camp Verde Marshals
Deputy Shawn Martinko	Camp Verde Marshals
Deputy Jacob Teague	Camp Verde Marshals
Officer Eric Hatchell	Chino Valley Police Department
Officer Vincent Schaan	Chino Valley Police Department
Officer Josh Fradette	Cottonwood Police Department
Officer Timothy Pierce	Cottonwood Police Department
Officer Nicole Horisi	Jerome Police Department
Officer Lang McGuire	Jerome Police Department
Officer Dan Mulleneaux	Phoenix Police Department
Officer Paul Clemens	Prescott Police Department
Officer Deana Marston	Prescott Police Department
Officer Dennis Martin	Prescott Police Department
Officer Ron Niederstadt	Prescott Police Department
Deputy Oscar Alvarez	Yavapai Co. Sheriff’s Office
Deputy Ryan Bair	Yavapai Co. Sheriff’s Office
Deputy Doug Brown	Yavapai Co. Sheriff’s Office
Deputy Ryan Goodell	Yavapai Co. Sheriff’s Office
Deputy Darrin Harper	Yavapai Co. Sheriff’s Office
Deputy Gerald D. McNally	Yavapai Co. Sheriff’s Office
Deputy William Pearson	Yavapai Co. Sheriff’s Office
Deputy Dan Raiss	Yavapai Co. Sheriff’s Office
Sergeant Bill Suttle	Yavapai Co. Sheriff’s Office
Deputy Jody Villalobos	Yavapai Co. Sheriff’s Office
Kim Abbott	
Michelle Spirk	Arizona DPS Crime Lab
Rita C. Dyas	Arizona DPS Crime Lab
Joe Slowinski	Arizona DPS Crime Lab
Robert Stephenson	Arizona DPS Crime Lab

PHOENIX POLICE DEPARTMENT

PHOENIX, AZ
MAY 16, 2003

Officer	Agency
Officer Dan Mulleneaux	Phoenix Police Department.
Officer Pete Smith	Gilbert Police Department
Sergeant Dick Studdard	Los Angeles Police Department (ret.)
Officer Adam Geremia	Phoenix Police Department
Officer Herbert A. Jacobs	Phoenix Police Department
Detective Jason Jahnke	Phoenix Police Department
Officer Darren Nielsen	Phoenix Police Department
Officer Jeffrey Riddle	Phoenix Police Department
Officer Michael T. Rogers	Phoenix Police Department
Sergeant Robert Smedes	Phoenix Police Department
Officer Tom Tardy	Phoenix Police Department
Officer Virgil Toland	Phoenix Police Department
Officer Chris Treadway	Phoenix Police Department
Officer Joe Geremia	Scottsdale Police Dept.
Michelle A. Spirk	Arizona DPS Crime Lab
Rita C. Dyas	Arizona DPS Crime Lab
Joe Slowinski	Arizona DPS Crime Lab

SAN ANTONIO POLICE DEPARTMENT

SAN ANTONIO, TX
JUNE 23-24, 2003

Officer	Agency
Albert Reeder	TEEX – Texas A&M Univ.
Cpl. Dennis Kelley	Alamo Heights Police Department
Officer David Boyd	Austin Police Department
Officer Ryan Herring	Austin Police Department
Officer Richard Bryan	Balcones Heights Police Department
Detective Mark Busbee	Boerne Police Department
Detective Wayne Lehman	Comal County Sheriff’s Office
Officer Joshua Bruegger	Pasadena Police Department
Officer Guadalupe Campbell	San Antonio Police Department
Officer Michael Field	San Antonio Police Department
Officer Scott Foulke	San Antonio Police Department
Officer Richard Long	San Antonio Police Department
Officer Michael Moore	San Antonio Police Department

Appendix VI

Characteristics of Alcohol Workshop Drinking Participants

California Highway Patrol
Sacramento, CA

July 29					
Participant	Gender	Age	Height (in)	Weight (lbs)	Peak BAC (g/dL)
1	F	28	59	115	.127
2	F	24	66	150	.160
3	M	34	68	154	.113
4	M	25	71	180	.103
5	M	29	72	185	.184
6	M	25	68	165	.090
7	M	33	69	150	.116
8	M	31	69	170	.097
July 30					
1	F	30	64	146	.083
2	M	35	77	225	.120
3	M	27	69	150	.098
4	F	29	66	141	.110
5	M	22	73	192	.089
6	M	24	72	190	.137
7	M	25	71	189	.126
8	M	32	72	165	.095

Phoenix Police Department
Phoenix, AZ

May 16					
Participant	Gender	Age	Height (in)	Weight (lbs)	Peak BAC (g/dL)
1	F	30	63	160	.132
2	M	49	70	210	.075
3	F	48	62	150	.075
4	F	29	62	145	.098
5	F	27	64	157	.091
6	M	29	70	180	.140
7	M	35	71	205	.098
8	F	27	64	106	.096
9	M	30	73	185	.083
10	M	31	70	215	.058

Northern Arizona Training Center
Prescott, AZ

May 2					
Participant	Gender	Age	Height (in)	Weight (lbs)	Peak BAC (g/dL)
1	F	28	64	200	.087
2	F	41	66	180	.099
3	F	29	60	125	.103
4	F	33	67	155	.123
5	F	28	65	150	.145
6	F	49	68	165	.104
7	F	27	64	140	.069
8	F	28	62	105	.128
9	F	29	63	140	.173
10	M	25	72	190	.039
11	M	27	69	190	.168
12	M	23	75	180	.064
13	F	45	65	130	.088
14	F	-	64	120	.128
15	F	25	66	200	.145
16	F	23	62	140	.147
17	F	38	64	125	.179
18	F	39	67	130	.148

San Antonio Police Department
San Antonio, TX

June 23					
Participant	Gender	Age	Height	Weight	Peak BAC (g/dL)
1	M	22	71	187	.099
2	M	22	66	168	.090
3	M	24	67	155	.122
4	M	38	65	163	.093
5	F	38	67	189	.027
6	F	37	64	185	.037
7	F	31	66	127	.121
8	F	34	63	197	.064
9	M	44	67	215	.097
10	M	34	70	207	.073
11	M	32	66	169	.102
12	F	36	64	145	.116
13	M	40		183	.109
June 24					
1	M	55	71	226	.097
2	F	27	68	185	.069
3	F	30	62	153	.106
4	F	26	62	168	.125
5	F	36		136	.025
6	F	37	64	216	.037
7	F	38	67	190	.100
8	F	25	62	112	.096

Appendix VII

REPORT OF THE
ANALYSIS OF FIELD DATA

Barbara Tabachnick, Ph.D.

I. Analytic Approach

For all analyses, the six indicators of officer-observed HGN (three signs for each eye) yielded a scale varying from zero (no indicators) to six (all three signs in both eyes). These HGN scores were recorded for each exam in all conditions (positions). The number of exams per participant was highly variable, ranging from to 5 to 46. Elapsed time was calculated as number of minutes since the first exam. Distributions of observed HGN scores were unacceptably negatively skewed (z = -23.01, $p < .001$) with about 70% of the responses in the highest category. The variable was dichotomized with the four lowest categories combined and the two highest combined. This produced a distribution suitable for parametric analysis.

Because of the differing number of exams for participants and the varying time intervals between exams, a multilevel modeling (MLM) strategy was taken to estimate means and standard errors. This strategy provided potential adjustment for individual differences (random intercepts for participants), location differences (random intercepts for locations), pre-exam BAC, and time of testing (i.e., lowering of BAC), as well as the interaction between condition and time of testing. Details of the modeling process are in Appendix VII.A. The means and standard errors were then used in the EquivTest software (Statistical Solutions), which develops tests of equivalences of pairs of means.

Equivalences between means for a pair of conditions were evaluated in EquivTest (Statistical Solutions) using a Schuirmann OST/TOST procedure in which one-tailed *nonequivalence* tests are applied to each of two null hypotheses: one for an upper bound of nonequivalence between means and another for a lower bound. A meaningful difference using a criterion of ± 20% (a conservative industry standard) of the HGN scores in the baseline condition was used for the Schuirmann tests. Rejection of the null hypothesis for these tests in both directions indicates that equivalence can be claimed. ***Non-rejection of either test is associated with nonequivalence.***

Two planned analyses included comparisons of standard position (standing) versus sitting and standard versus lying down.

II. Variations in Participants' Positions

Standard (standing) versus sitting

A two-level [MJ1]model (examinations at the first level and individuals at the second level) showed significant individual differences in dichotomized HGN, Wald $z = 4.03$, $p < .001$. The intraclass correlation of .43 indicates that 43% of the variability in HGN is explained by differences among participants and underscores the need to take those differences into account as a random effect in modeling variability in HGN.

The best-fitting model of the effect of standing versus sitting on dichotomized HGN included pre-exam BAC as a covariate as well as individual differences as a random effect. Using this model, the 90% confidence interval for the differences in adjusted, dichotomized HGN scores between the two conditions was -0.031 to 0.177. Both the upper and lower bound nonequivalence tests yielded rejection of the null hypotheses, so that a claim of equivalence is supported.

Standard (standing) versus lying down
A two-level [MJ2]model for this comparison showed significant individual differences in dichotomized HGN, Wald $z = 14.47$, $p < .001$ with an intraclass correlation of .44. The best-fitting model of the effect of standing versus lying down on dichotomized HGN included pre-exam BAC as a covariate as well as individual differences as a random effect. Using this model, the 90% confidence interval for the differences in adjusted, dichotomized HGN scores between the two conditions was -0.043 to 0.151. Both the upper and lower bound nonequivalence tests yielded rejection of the null hypotheses, so that a claim of equivalence is supported.

Using this model, the 90% confidence interval for the differences in adjusted, dichotomized HGN scores between the two conditions was -0.043 to 0.151. Both the upper and lower bound nonequivalence tests yielded rejection of the null hypotheses, so that a claim of equivalence is supported.

Appendix VII.A

Separate multilevel models were developed for the two planned comparisons: standing versus sitting and standing versus lying down. Three-level models were first run to determine the need for taking into account individual and location differences. In these models, examinations were the first-level units, individual participants the second-level units, and locations the third-level units. For both comparisons, location differences were too minimal to require consideration as a random effect. For sitting versus standing, $\rho = .01$, Wald $z = 0.22$, $p = .828$. For standing versus lying down, $\rho = .04$, Wald $z = 0.603$, $p = .546$. Therefore, remaining models were analyzed with two levels: examinations and participants. All models were run with maximum likelihood estimation to permit comparisons among them.

Position effects, pre-examination BAC, elapsed time since first exam, and the interaction between elapsed time and position were added to the two-level models for both comparisons, clearly improving them above the null models. However, the latter two effects were found to be non-significant as covariates for predicting dichotomized HGN from position. That is, they failed to provide adjustment of dichotomized HGN in testing the effect of position. Therefore, the final model used for determining adjusted, dichotomized HGN means and standard errors was a model that included position as an independent variable and BAC as a covariate as well as individual differences as a random effect. Tables 1 and 2 summarize the four models for the two planned comparisons. In these tables, a significant difference between models indicates that the more complex model is reliably better than the less complex one.

Table 1 Comparison of Nested Two- and Three-Level Models for Dichotomized HGN, Standing versus Sitting Positions

Model	LR	df	χ^2 difference test
Three-level model			
Model 1 Three-level intercepts only (random effects)	519.1	4	
Two-level models			
Model 2 Two-level intercepts only (random effect)	519.1	3	M2- M3 = 0.0
Model 3 Two-level model with position and BAC (both fixed effects): Final Model	491.0	5	M2 - M3 = 28.1*
Model 4 Two-level model with position, BAC, elapsed time, and time by position interaction (all fixed effects)	488.3	7	M3 - M4 = 2.7

Table 2 Comparison of Nested Two- and Three-Level Models for Dichotomized HGN, Standing versus Lying Down Positions

Model	LR	df	χ^2 difference test
Three-level model			
Model 1 Three-level intercepts only (random effects)	390.5	4	
Two-level models			
Model 2 Two-level intercepts only (random effect)	391.2	3	M2 - M1=0.7
Model 3 Two-level model with position and BAC (both fixed effects): Final Model	365.1	5	M2 - M3 = 26.1*
Model 4 Two-level model with position, BAC, elapsed time, and time by position interaction (all fixed effects)	363.9	7	M4 - M3 = 0.2

*$p < .01$

DOT HS 810 831
September 2007

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

NHTSA
www.nhtsa.gov