December 1981
Final Report

DOT HS-806-211

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

An Assessment of Behavioral Tests to Detect Impaired Drivers

K. J. Snapper
D. A. Seaver
J. P. Schwartz

Decision Science Consortium, Inc.
7700 Leesburg Pike, Suite 421
Falls Church, Virginia 22043

Contract No. DTNH 22-80-C-07565
Contract Amount $67,766

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT-HS-806 211		

4. Title and Subtitle	5. Report Date
AN ASSESSMENT OF BEHAVIORAL TESTS TO DETECT IMPAIRED DRIVERS	December 1981
	6. Performing Organization Code
7. Author's) K. J. Snapper, D. A. Seaver, J. P. Schwartz*	8. Performing Organization Report No. 81-11
9. Performing Organization Name and Address Decision Science Consortium, Inc. 7700 Leesburg Pike, Suite 421 Falls Church, Virginia 22043	10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTNH22-80-C-07565
12. Sponsoring Agency Name and Address U.S. Department of Transportation National Highway and Traffic Safety Administration Problem Behavior Research Division Washington, D.C. 20590	13. Type of Report and Period Covered Final Report August 1980 - December 1981 14. Sponsoring Agency Code

15. Supplementary Notes

*With a contribution by Elliot R. Levine, Esq.

16. Abstract

This study was conducted to assess the feasibility and practicality of roadside behavioral tests for driving impairment. Such tests are seen as complementary to chemical tests of blood drug concentration. Methodological issues in the use of behavioral tests are identified and discussed focusing specifically on the validity of behavioral measures. It is argued that if validity is defined as predicting accidents, it will always be low for any behavioral test as well as for chemical tests because of the inherent unpredictability of accidents. The rationale for using behavioral tests, particularly a battery of behavioral tests, to detect driving impairment is, then, that such tests can measure the level of behavioral skills involved in driving. In order to evaluate various behavioral tests, a large literature was reviewed to identify potential tests and any previous studies of them. Criteria for evaluating the tests were developed, including validity, reliability, ease of operational use, safety, and diagnosticity. Each of 54 tests were scored on each criterion, and criteria were weighted by their relative importance from the perspectives of researchers and users to produce an overall index of the feasibility of each test. Tests that were rated highly include visual acuity, finger-to-finger, finger-to-nose, and romberg (and modified romberg). An analysis was also performed of legal issues relating to the use of roadside behavioral tests. Such tests appear to have useful status as evidence (primarily circumstantial), and probably can be used without a need for expert witnesses.

17. Key Words	18. Distribution Statement
Legal Issues; Behavioral Tests; Driving Impairment; Human Performance; Drug Effects; Multiattribute Utility Theory; Test Validity; Test Reliability; Test Safety; Operational Use of Test	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	122	

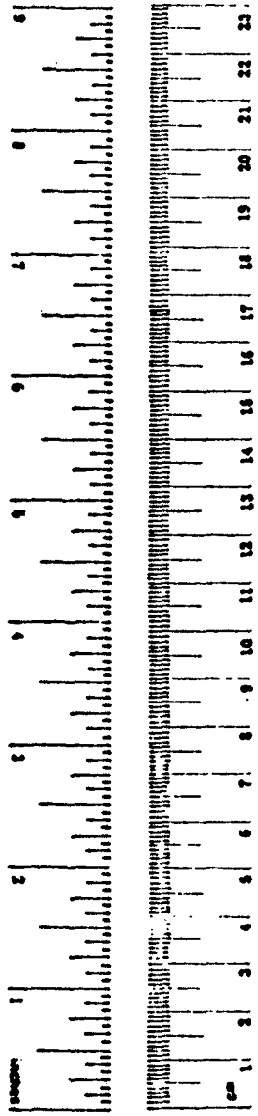
Form DOT F 1700.7 (8-72) Reproduction of form and completed page is authorized

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

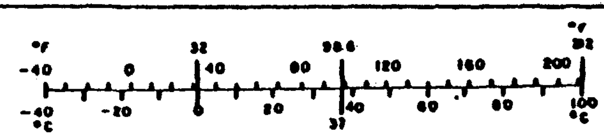
Symbol	When You Know	Multiply by	To Find	Symbol
	LENGTH			
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
	AREA			
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
	MASS (weight)			
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
	VOLUME			
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
	TEMPERATURE (exact)			
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
	LENGTH			
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
	AREA			
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
	MASS (weight)			
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
	VOLUME			
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
	TEMPERATURE (exact)			
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



METRIC CONVERSION FACTORS

TABLE OF CONTENTS

TABLE OF CONTENTS
(continued)

1.0 INTRODUCTION AND BACKGROUND

Drivers who are unable to operate their vehicles safely should not drive, but there is considerable controversy about how best to detect driver impairment. Two basically different approaches have been suggested: (1) development of chemical tests to detect the presence of impairing agents, and (2) development of behavioral tests to detect impairment. These approaches may be complementary, but neither approach has been fully developed or tested. This report assesses the present state and potential of behavioral testing methods, for the purpose of recommending whether further development is justifiable, based on the present state of the art. Throughout this report as "impairment" is discussed, it is important to keep in mind that we are not referring to impairment in general, but rather are discussing impairment as it specifically relates to driving.

1.1 Legislative Mandate for Study

Alcohol has long been identified as a risk-factor in traffic accidents. Among other studies, Borkenstein, Crowther, Shumate, Ziel, and Zylman (1964) have indicated that elevated blood-alcohol concentration (BAC) increases the likelihood of an accident, but the nature of the relationship between BAC and accidents is not agreed upon (e.g., Zylman, 1972, 1974). While there has been emphasis on alcohol and accident prevention--possibly because of the widespread use of alcohol and its easy detection in the driver--it is known that other drugs and conditions can impair driving.

The incidence of use and exposure rates are generally unknown, and very difficult to estimate, for drugs other than alcohol (e.g., see Donelson, Marks, Jones, and Joscelyn, 1980). That is, for drugs other than alcohol, there is no equivalent of the Borkenstein et al. study, which provides generally accepted estimates of utilization and risk levels. Drugs other than alcohol are also much harder to detect chemically, since compared to other drugs alcohol appears in extremely high concentrations in breath, blood, and tissue samples. It may also be harder for the law enforcement officer to detect other drugs using behavioral cues, since alcohol intoxication is a familar and easily recognizable phenomenon, and because there is no one behavioral syndrome for other drugs. For these and other reasons, the drinking-driver problem has been defined much more clearly than the problem of the impaired driver more generally.

Nevertheless, there has been increasing concern about drugs in general, and problems of detecting the impaired driver. Estimates of both legal and illegal drug consumption suggest that there must be significant exposure, though epidemiologic estimates vary widely (e.g., Sharma, 1976).

The concern with highway safety and drugs is expressed in Section 212 of Title II of the Surface Transportation Act of 1978 (known as the Highway Safety Act of 1978). The Act requires a report to the Congress on efforts to detect and prevent drug use of motor vehicle operators, specifically:

> Such report shall include, but not be limited to, information concerning the frequency of marijuana and drug use by motor vehicle operators, capabilities of law enforcement officers to detect the use of marijuana and drugs by motor vehicle operators, and a description of Federal and State projects undertaken into methods of detection and prevention. The report shall include the Secretary's recommendations on the need for legislation and specific programs aimed at reducing marijuana and other drug use by motor vehicle operators.

1.2 Technical Problems in Developing Tests of Impairment

The problem of detecting drug-induced impairment has typically been conceptualized as a measurement problem. Over the past several years, this has led to two rather distinct lines of inquiry. One line has sought to measure the presence of an impairing agent in the body. Currently, alcohol is the only drug for which accurate and generally accepted quantitation techniques are widely used.

The second line of inquiry has sought behavioral measures of impairment. One argument for behavioral tests is intuitive: a behavioral test should be better than a chemical test for measuring or predicting impairment because impairment is a behavioral condition. Although there is a very large literature on impairing agents and performance on a variety of tasks, there is no generally accepted behavioral task or battery of tasks for measuring or predicting impairment, though we and others have long argued the presumed but unproven merits of this approach (Edwards, Goodman, and Snapper, 1971; Kaplan, Lathrop, and Edwards, 1976; Edwards, Lathrop, Seaver, Seghers, and Lehman, 1973).

In its report to the Congress, NHTSA (1979) described both the chemical test and behavioral measure approaches to detecting and preventing drug-induced impairment and for enhancing enforcement efforts. That report identified a number of gaps in the current state of the art. Below are our interpretations and elaborations of some of the points raised in that report.

1.2.1 Non-invasive chemical tests. Practical as well as technical constraints currently limit the ability to quantify concentrations of drugs other than alcohol. Blood is the specimen of choice for quantitation purposes, though urine or saliva may be used for screening rather than quantitation purposes; and accuracy in quantitation requires costly equipment and considerable care and expertise in handling blood samples and equipment (Joscelyn and Donelson, 1980a). There is currently no inexpensive, practical but accurate equivalent of the non-invasive breath test for alcohol.

1.2.2 Ambiguity of blood drug concentration analyses. Joscelyn and Donelson (1980a) argue that interpretation of blood drug concentration analyses (BDC) "...depends on prior knowledge of what the analytic results means in terms of driver impairment." Joscelyn, Jones, Maickel, and Donelson (1979) argue that there have been few efforts to relate BDCs to driving-related skills for any drugs except alcohol. NHTSA (1979), among several other reports, describes the methodological limitations in those studies that have been done.

Whether or not a reasonably orderly relationship between BDC and impairment will ever emerge is of course an empirical question, whose verification would require extensive (and expensive) future research. Moreover, as Joscelyn and Donelson indicate, the use of multiple drugs and the interactions among them are "an increasingly frequent occurrence... [and] often present even greater problems for interpretation." For example, alcohol and barbiturates may produce a greater risk together than either does by itself (Sharma, 1976; Smith, 1966).

1.2.3 Defining impairment threshold in terms of BDC. As indicated above, whether or not a program of investigating BDC levels and driving impairment would produce a reasonably orderly relationship is an empirical question. The optimistic argument is that such an orderly relationship would emerge, and that it should be possible to establish (as for alcohol) concentrations that can be defined as comprising the threshold for impairment. The pessimistic view is that alcohol represents a special case--often being used alone, and in quite massive, easily detectable quantities--and that such orderly relationships would not emerge for other drugs. Moreover, it might be argued, the use of multiple drugs might produce impairment well below the threshold established for any single drug, so a single threshold cannot be established relating BDC to impairment.

Another pessimistic argument against development of the use of BDCs to define impairment is that, even in the case of alcohol, individual differences among individual drivers are so large that any BDC-defined criterion of impairment must of necessity be unreliable and introduce either large Type 1 or Type 2 errors. For example, Snapper (1973) showed large, systematic differences between light and moderate drinkers; and Burns and Moskowitz' data (1977) suggest that .10 blood-alcohol concentration (BAC) is at best a fuzzy criterion of impairment. As a consequence, in order to ensure that "not too many" unimpaired drivers are charged, the BAC criterion must be set high, if not by statute then in practice by the law enforcement officer. This argument, pushed to its logical extreme, holds that BAC and/or BDC are in principle inappropriate criteria and that some external, non-chemical standard for defining impairment is more appropriate.

1.2.4 Specificity of behavioral tests. Critics of behavioral test approaches could well argue that the reason there is an imperfect relationship between BAC and performance is because the behavioral tests themselves provide imperfect measures. A critic of behavioral tests could argue that BAC or BDC is more strongly related than behavioral measures to "true" driving skill; and, given the extremely poor definition of the driving task, the argument in fact seems moot based on existing empirical evidence.

In its most relevant version, this argument would postulate that simple behavioral tests--of the sort that could be used roadside or in the police station --are not specific to the driving task. If such a test were to disagree with a chemical test, the argument continues, it is likely to be because the behavioral test is measuring something that has nothing in particular to do with driving.

1.2.5 Practicability. Both chemical and behavioral testing procedures may be questioned on grounds of their practicability. It may be argued that the costs of instrumentation are high, that extensive training is required, that law enforcement officers will object to the testing procedures, or that the courts will require expert witnesses to testify to results. The concern would be that, even if valid chemical and/or behavioral testing procedures were to be developed they could not be deployed because of practical issues.

1.3 Scope of Present Study: Assessment of Behavioral Test Procedures

It is not the purpose of this study to compare chemical test procedures against behavioral test procedures. It is clear that, on practical grounds at least, both approaches require considerably fuller development. Indeed, it may be that neither will prove adequate, in terms of yielding quantitatively meaningful results, using practicable procedures. In order to describe the purpose of the present study better, it is useful briefly to compare and contrast the problems of developing chemical versus behavioral tests.

The problems in developing chemical versus behavioral tests are quite different. A good chemical test procedure yields accurate and reliable quantitative results in regard to a well defined variable: the actual concentration of some substance in the blood. The critical question appears to be one of adequate training, equipment cost and reliability, invasiveness of techniques required to obtain body-fluid samples, and of course the problem of translating chemical analysis results into assessments about degree of impairment.

The last issue, assessing degree of impairment, is in contrast where the behavioral approach begins. Whereas the chemical approach attempts to avoid defining precisely what is meant by impairment, behavioral approaches run into this problem immediately. The matter of defining and assessing impairment raises some extremely complex and slippery conceptual and methodological measurement questions; compared to these, instrumentation issues are much less critical (though still non-trivial).

The purpose of this study is to assess the feasibility of behavioral tests of impairment. The factors we regard as important are defining operationally what is meant by "impairment;" and determining whether there are valid measures that can be practically administered. In contrast, we will be less concerned about specific instrumentation problems, or about recommending a particular battery of tests for fuller future development.

We should note explicitly that we are not concerned with comparatively evaluating chemical versus behavioral tests, with an eye towards influencing future decisions about which (if either) is further developed and deployed. Our opinion is that both types of tests can be used, though probably in different ways. We see the two approaches as being complementary, not competitive. We will make explicit our bias, and that is that behavioral tests, in principle, offer more valid and justifiable measures of impairment then chemical tests do or ever can. However, the crucial question is the extent to which either approach can be more fully developed, and if either can offer practical assistance in regard to detection of impaired drivers and law enforcement. This report describes our conclusions and supporting evidence regarding further pursuit of behavioral testing techniques.

2.0 IMPAIRMENT MEASURES: METHODOLOGICAL ISSUES IN INDEX CONSTRUCTION, EVALUATION, AND VALIDATION

2.1 The Criterion Problem

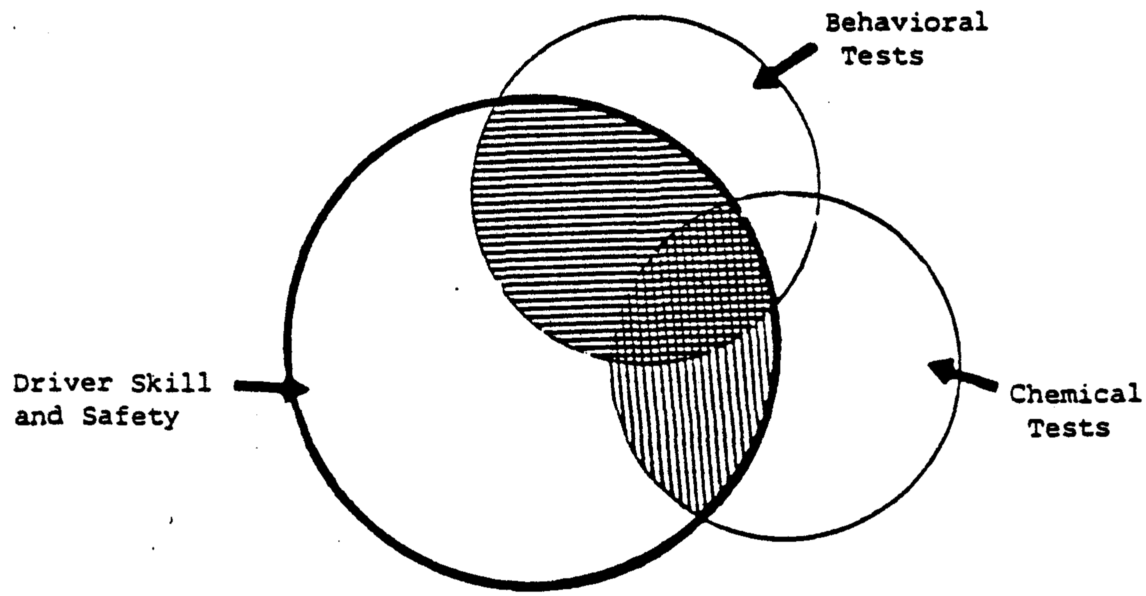
As suggested in Section 1.0 of this report, detection of the impaired driver is fundamentally a measurement problem. However, what is one trying to measure? Although there are widely differing opinions about how one should test for impairment, there is a reasonable consensus that impairment (like driver competence) should be defined in terms of driver skill and safety. We will assume, therefore, that "Driver Skill and Safety" is the underlying concept we are trying to assess, regardless of whether we use chemical or behavioral tests. Throughout the discussion here, it is important to keep in mind the role of existing data. Very few data relate directly to accident risk; what data exist, and they are sparse, related primarily to impairment of performance.

"Driver Skill and Safety" is what Thorndike (1949) called an ultimate criterion. That is, if we had a direct measure of Driver Skill and Safety, we would not need indirect and partially irrelevant chemical or behavioral tests. Moreover, if such a direct measure of Driver Skill and Safety existed, we would be able to determine which behavioral tests were good measures of impairment, and we would be able to determine how favorably chemical tests compared to behavioral tests.

With both chemical and behavioral tests, we are using imperfect measures of Driver Skill and Safety, as illustrated in Figure 1. (This figure, and those that follow are meant only to illustrate conceptual relationships. They should not be interpreted literally as precise representations of the relationships.) The concept of Driver Skill and Safety is rich, spanning virtually all facets of human performance, perception, and cognition. It involves both the internal or endogenous properties of the driver, but also the interaction between the driver and external or exogenous factors (such as the vehicle itself, other vehicles, ambient conditions, etc.). Many impairing agents or conditions affect only certain aspects of Driver Skill and Safety, so chemical tests can, in principle, assess certain aspects of Driver Skill and Safety. A similar

FIGURE 1

BEHAVIORAL AND CHEMICAL TESTS AS
IMPERFECT MEASURES OF DRIVER SKILL AND SAFETY



point can be made for behavioral tests, which may reflect some very specific aspects of Driver Skill and Safety. And neither behavioral or chemical tests are perfect measures: each will be partially irrelevant either conceptually or in terms of introducing statistical error variance.

2.2 Operational Criteria

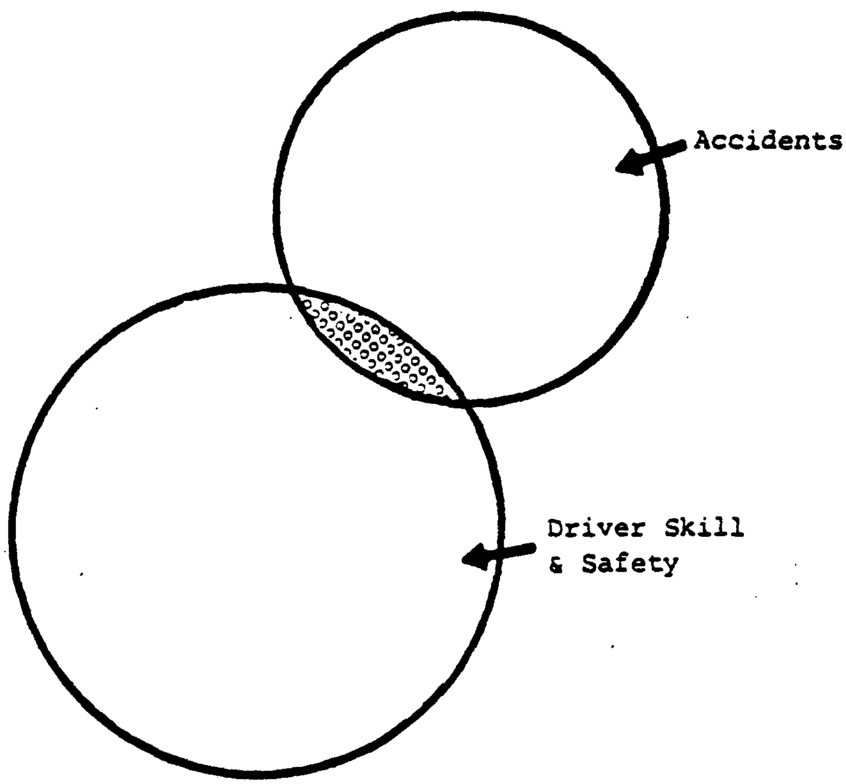
Clearly, the concept of Driver Skill and Safety exists only as an idealized construct. There is at present no way to operationalize the concept in order to use it directly for measuring impairment. Indeed, while McKnight and Adams (1970) have delineated driving behaviors (typically these are tasks performed by the driver), it is unclear what human skills are required to execute those driving behaviors successfully. Joscelyn and Donelson (p. 16, 1980b) indicate the problem of determining "the relevance of test measures to the driving task" and that this is largely a judgmental matter, given the very limited empirical information.

Several surrogate or proxy criteria can be identified as practical, operational criteria of driving skill, safety, and impairment. While probably no one would suggest that any of these "operational criteria" in any sense captures all of what is meant by Driver Skill and Safety, one might plausibly argue for them on grounds of practicality. Before explaining the approach we propose for development of behavioral measures of impairment, it is useful to introduce several quantitative measurement issues by discussing four possible operational criteria of impairment.

2.2.1 Traffic accidents as an operational criterion of impairment. The argument for this criterion is that unskilled or unsafe driving will produce accidents and, therefore, a measure of impairment should be predictive of accidents. This is illustrated in Figure 2 by the overlap between Driver Skill and Safety and Accidents. This overlap indicates the "relevance" (See Thorndike, 1949, or Snapper, O'Connor, and Einhorn, 1974 for a fuller discussion of this concept) of Accidents as an operational criterion for Driver Skill and Safety. ("Relevance" can refer either to a conceptual or logical subsuming of one criterion by another; or it can refer to a statistical correlation. We will

FIGURE 2

ACCIDENTS AS AN OPERATIONAL CRITERION
FOR DRIVER SKILL AND SAFETY



distinguish between these two senses of "relevance" when there is risk of misunderstanding.)

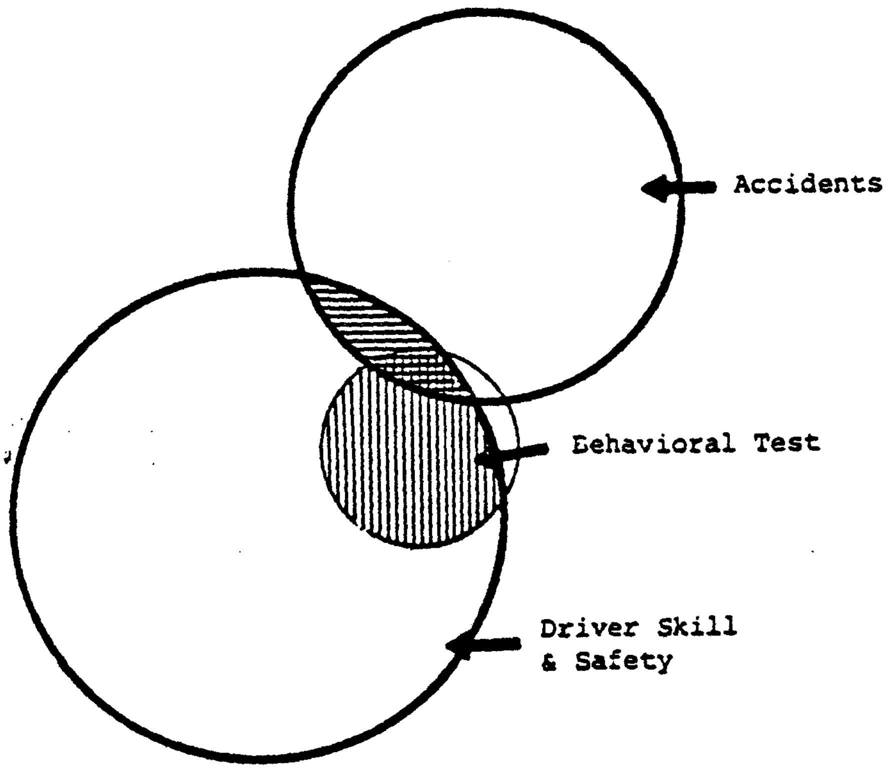
The unshaded area indicates the irrelevance of Accidents as an operational measure. An argument can be made that Accidents is a quite poor operational criterion because there is a very weak relation between driving performance and accidents. For instance, even at high BACs it is improbable that a driver will have an accident: accidents are unpredictable, in an absolute sense. (This is not to deny, however, that the relative probability of an accident increases with BAC; we are arguing only that absolute probabilities--and therefore predictability--remain quite low.) Goldstein (1961) argues that the inherent predictability of accidents is very low. Accidents have very low autocorrelation (as assessed by correlations from one year to the next) of around r=.10, implying that no other measure (chemical, behavioral, or other) is likely to predict them well. Of course, one reason accidents are not highly related to driver performance is because accidents stem as much from exogenous conditions and the driving context as from the condition of the driver. Thus, while Accidents may be conceptually related to Driving Skill and Safety, the inherently poor statistical predictability of Accidents flaws it deeply as an operational criterion of impairment, as a practical matter.

2.2.2 A behavioral test as an operational criterion of impairment. Several simple behavioral tests (like a test of reaction time) are candidates as measures of impairment. We discuss here whether any such simple behavioral test can serve as an operational criterion. In Figure 3 we assume that the test chosen is relevant to Driver Skill and Safety. However, the relationship will be weak, since a given behavioral test is unlikely to be relevant to more than a very limited range of driving skills. This means, for example, that it would be difficult to validate a given behavioral test using behind-the-wheel studies; it would only be by coincidence that whatever skill is measured by the particular test would be revealed in the driving task.

In Figure 2 we illustrated the relative weak relationship between Accidents and Driving Skill and Safety. Partly because of this low predictive relationship, a behavioral test is unlikely to have a measurable statistical relationship

FIGURE 3

BEHAVIORAL TEST AS AN OPERATIONAL CRITERION
FOR DRIVER SKILL AND SAFETY AND ITS RELATIONSHIP TO ACCIDENTS



with Accidents. Goldstein's (1961) findings provide empirical support for this argument. Thus, even valid behavioral tests are likely to be poor predictors either of actual behind-the-wheel driving (because of the complexity of the driving task and the fact that a single test can be relevant only to a small subset of driving skills) or of accidents.

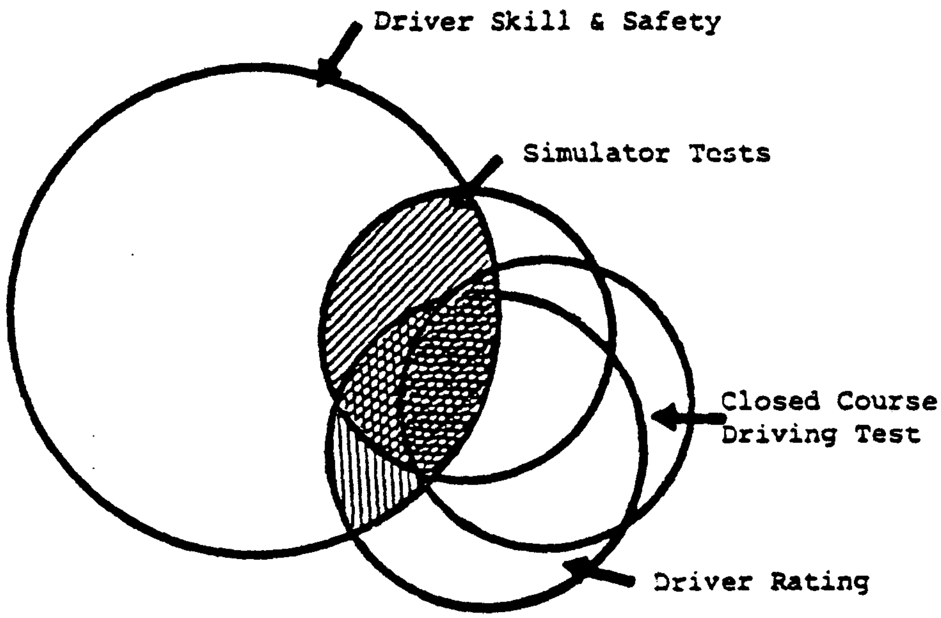
A slightly more general and slightly overstated version of this argument is that a simple behavioral test is unlikely to be correlated with much of anything, and would be extraordinarily difficult to validate using actual driving tasks, simulators, closed course driving tasks, or accidents as the criterion of validity.

2.2.3 Driving-like tasks as an operational criterion of impairment. Closed-course driving tasks, actual behind-the-wheel driving scored by an observer, and simulators have been used to assess "driving" performance. This type of test procedure has been proposed as an operational criterion. As shown in Figure 4, Driving Tasks are likely to be partially irrelevant because they are affected by exogenous contextual factors not encountered in "real" driving. It may well be that these types of "driving" task-scores correlate with one another partially because of factors that are exogenous to "real" driving. For example, Moskowitz (1975) examined whether simulator and other studies on "driving tasks" produced correlated results; any such correlations could be due to factors unrelated to Driving Skill and Safety, and there is no statistical method for proving relevance and validity.

More troublesome is the fact that "driving-like" tasks are likely to involve only a comparatively small subset of what is implied by Driver Skill and Safety. This is perhaps most apparent in closed-course driving, where the range of situations that can arise is impoverished compared to what a typical driver must routinely handle in real driving. This means that, even if great efforts are made to make driving-like tasks more realistic, there is likely to be an inherently low degree of relevance to Driver Skill and Safety overall. Of course, as a practical matter, driving-like tasks could not be administered at roadside.

FIGURE 4

RELATIONSHIP BETWEEN DRIVING TASKS AND DRIVER SKILL AND SAFETY



2.2.4 BDC as an operational criterion of impairment. Blood drug concentrations (especially BAC) are often regarded as operational criteria. This is established by fiat for BAC in "per se" laws, which define impairment in terms of the BAC. An argument can be made, however, that BDC is from a measurement perspective not particularly as good as an operational criterion.

The relationship between BDC and Driver Skill and Safety is shown in Figure 5, which also displays Accidents. Accidents as an operational criterion is displayed here to underscore a point made earlier. Neither BDC nor any other measure is likely to predict accidents well; and those skills to which BDC is related may or may not be those that are implicated in accident generation. Thus, BDC is unlikely to be strongly related in an absolute sense to accidents.

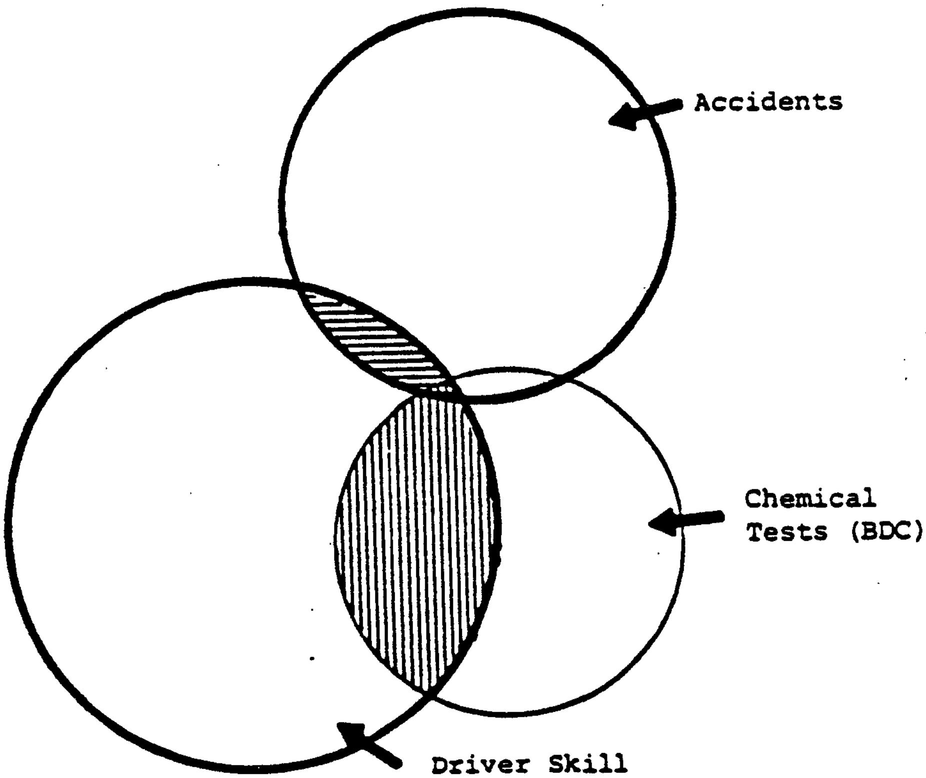
How strong is the relationship between BDC and Driver Skill and Safety? The answer depends of course upon what drug is involved, driver populations, and the drug concentration used. Clearly at extremes, say BACs above .30, or when the subject has passed out, all functions will be impaired. The fact does not prove that BDC is a satisfactory measure or operational criterion for Driver Skill and Safety.

At moderate BDCs, and in the "everyday" population of drivers, there are arguments that BDC is not likely to be satisfactory as an operational criterion. We will consider BAC as an example. First, there are very large individual differences among people, due to tolerance and habituation, and due to the fact the certain overlearned skills may be relatively impervious to drug-induced impairment (e.g., Snapper, 1973; Wallgren and Barry, 1970, p. 353; Zabik, 1977). Second, there is the problem of multiple impairing agents; for example, multiple drugs may cause impairment when no one of them is elevated above a critical threshold. Third, a given BDC will reflect just some of the skills that may be impaired; the marginal driver, the fatigued driver, or the driver using multiple drugs may experience forms of impairment that would not typically be indicated by a given drug concentration.

Of course, the above problems with BDCs have been recognized. And the defense has tended to be an argument that, in potential arrest populations, BDCs are so

FIGURE 5

BLOOD DRUG CONCENTRATION AS AN OPERATIONAL CRITERION FOR DRIVER SKILL AND SAFETY AND ITS RELATIONSHIP TO ACCIDENTS



elevated that the driver "must" be impaired. And, indeed, if one restricts BDCs only to high levels of concentration (e.g., BACs above .15) the measurement properties of BDCs improve. That approach leaves open the matter of assessing impairment in populations whose consumption of impairing agents has been more modest.

2.3 Some Tenets Regarding Development of Measures of Impairment

The foregoing discussion suggests that there are definite limits to any type of measure or operational criterion of impairment. This is not to imply, however, that behavioral tests of driving impairment are not useful. But it is important to understand the limitations on the measurement of impairment, in order to understand the role that behavioral tests can play in detecting driving impairment. Thus, here we first point out these limitations before setting forth an argument for behavior tests of driving impairment.

First, both chemical and behavioral measures (including results of driving-like tasks) can capture only a relatively small part of what is generally meant by Driver Skill and Safety. This fact means that there will be an inherently low correlational or other relationship between such operational criteria and Driver Skill and Safety in actual driving. Any of the foregoing measurement approaches should have a positive relationship with what one means by Driver Skill and Safety, but the conceptual and statistical relationships are not likely to be strong enough to be very satisfying. One must therefore recognize that there are probably inherent limits to the goodness of any practical operational criterion, in relation to Driver Skill and Safety.

Second, a measure of impairment may be valid, conceptually, it may have the greatest feasible predictive power, and it may still be a poor and unsatisfying predictor of accidents. Relative probabilities of an accident may increase, but in an absolute sense predictability will always be low. Such low predictability is not something that can be readily overcome, since accidents are inherently of very low predictability.

Third, BDCs (especially BAC) may be associated with increasing relative probability of an accident (as shown by Perrine, Waller, and Harris, 1971; Borkenstein, Crowther, Shumate, Ziel, and Zylman, 1964; Farris, Malone, and Lilliefors, 1976; McCarroll and Haddon, 1962; and others, in the case of BAC). This increase in relative probability is consistent with the low absolute predictability of accidents. Moreover, this result is consistent with past observations about inconsistencies across individuals and circumstances in dose-response relationships involving driving and other tasks.

Fourth, a given behavioral test may be a valid measure of Driver Skill and Safety, but is likely to be relevant to only a vary narrow aspect thereof. All behavioral tests will, like chemical tests, be severely limited in their ability in an absolute sense to predict accidents.

Fifth, validation of any type of measure or operational criterion of impairment will be difficult. Even the best feasible measure or criterion may be expected to be extremely difficult to validate, if by validation one means prediction of accidents and/or <u>real</u> driver skill and safety.

Sixth, validation of a measure or criterion of impairment can take several forms. One is verifying that the measure is relevant to Driver Skill and Safety--a task that is more judgmental than empirical. A second is statistical validation. Statistical validation, irrespective of whether a chemical or behavioral test is used, should produce for a valid measure or criterion a consistent but relatively weak pattern of correlations with: simulator data, closed-course performance data, driver-rating data, and such measures of "actual" driver performance as may be obtained; and with accident statistics.

Seventh, there is no incompatibility between behavioral measures and chemical tests: the two could be used simultaneously to assess Driver Skill and Safety. An example of this is offered by Kataja, Penttila, and Tenhu (1975), though there are others. Issues for determining use of behavioral versus chemical tests could depend more upon practical considerations than upon any fundamental empirical advantage that one has over the other.

Perhaps our major conclusion from reviewing the foregoing measurement issues is that even a highly valid measure or criterion of impairment will generate empirical validation results that many will find disappointing. For example, some people have argued that such a measure should predict accidents, and our suspicion is that they greatly overestimate the degree of predictability that is attainable, irrespective of whether they personally favor behavioral or chemical test approaches. Validity, in regard to behavioral tests, is probably best defined not as predictive validity but as construct validity. We will discuss the implications of this conclusion for the development of behavioral tests of impairment more fully in Section 2.4, below.

2.4 Procedures for the Development of a Behavioral Index of Impairment

What are the desirable properties of a behavioral index of impairment? In this section, we will discuss the properties of such an index qua operational criterion of impairment. When we refer to an "index of impairment" we will be referring to the aggregated scores of a battery of behavioral tests.

We do not believe that predictive validity of accidents, or of "actual" driving behavior, should be invoked as the touchstone of whether a particular behavioral index is valid or not. The reason for this is that neither is likely to be adequately predicted by any type of impairment measure, behavioral or chemical.

We believe that "impairment" should be defined as a significant reduction in the capability of a driver to execute skillfully and safely driving tasks that potentially may arise. Such an index would assess perceptual, motor skill, and other factors that are relevent to Driver Skill and Safety. Such relevance would primarily be determined judgmentally, both because of the general inadequacy of the empirical data base and because Driver Skill and Safety is an ill-defined construct. (An example may help illustrate this point. Ideally, it would be possible to identify certain skills necessary for safe driving, and to identify some direct measures of these skills. Any reduction in these measures would be regarded, therefore, as evidence that the driver is impaired.)

This definition of impairment as reduced capability is compatible with observations made by Joscelyn and Donelson (1980b). Paraphrasing them (p. 21), they argue that past analyses of the driving task have not focused on the problem of measuring drug-induced impairment. For instance, some analyses have analyzed the driving task at a level appropriate to driver education training. Impairing agents act, however, on the behavioral capabilities of the driver, rather than directly on any particular behavior. As one of their participants stated the point, impairing agents do not act directly on making left-hand turns. As we are arguing, impairment therefore is most directly expressed in terms of reduced capabilities, and assessment of such capabilities is the most appropriate way to measure it. Clearly, any such behavioral index of impairment should be predictive of accidents in validation studies. But we suggest that such predictability should be expected only in a relative sense, not in an absolute sense. That is, one should anticipate that higher index scores would be associated with a higher relative probability of driving errors or accidents.

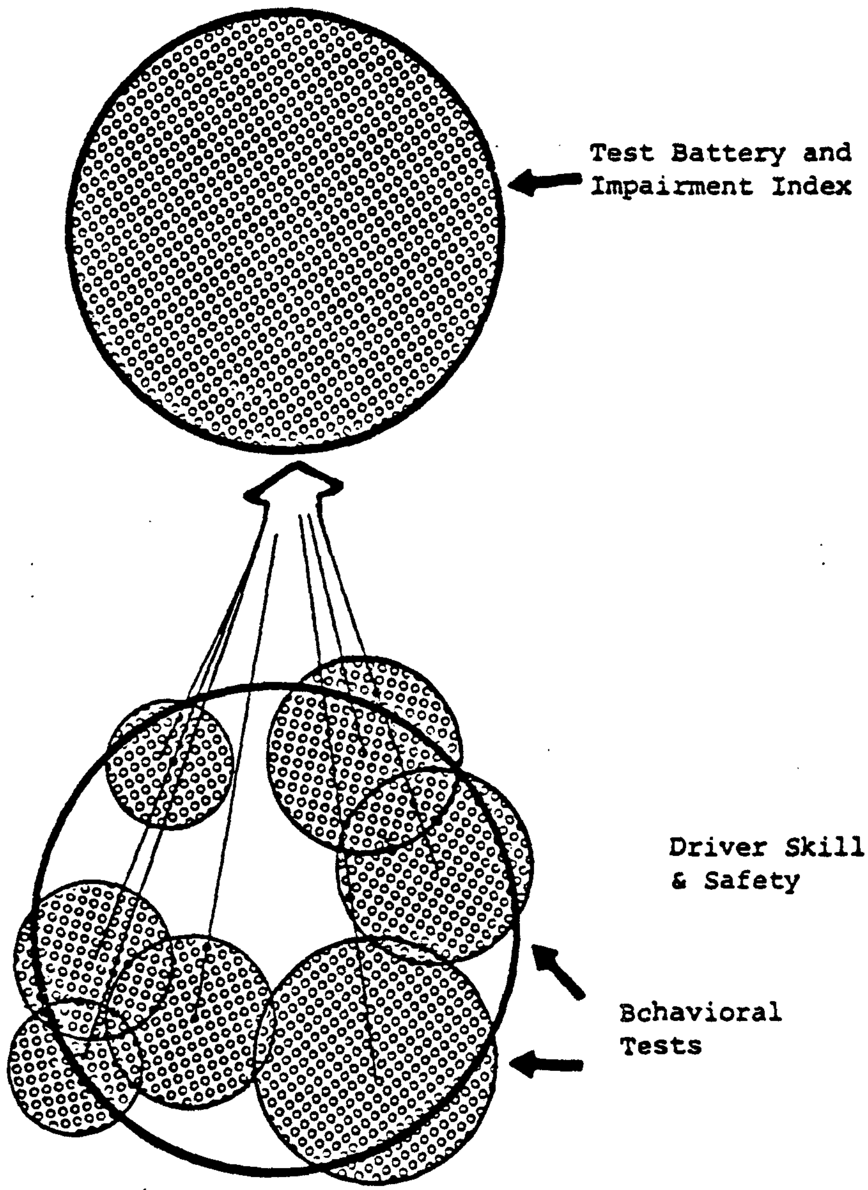
Earlier in this section we considered some of the problems with individual behavioral tests used as operational criteria of impairment. We now consider issues involved in developing a battery of behavioral tests, and their use as a criterion of impairment.

One problem, solved by use of a battery of tests, is the fact that any one test will be relevant to only a very limited part of Driver Skill and Safety. A battery of tests should be generated so that it is relevant to a representative range of driving skills, as indicated by Figure 6. Tests must be selected on largely judgmental grounds to ensure that they assess a representative range of driving skills, as indicated by the dispersion of tests over Driver Skill and Safety. Although there may be empirical data concerning particular behavioral tests, the process of relating tests to Driver Skill and Safety is nevertheless judgmental.

It is not necessary that these tests be shown to be particularly sensitive to low doses or levels of drugs or other impairers: what is necessary is that the test be a valid measure of Driver Skill and Safety. Ideally, experts would agree that inability to perform on a given test implies that driver capability has been diminished, or better yet, a relationship between impairment and accident risk could be established. By definition, then, the test would offer a

FIGURE 6

A BATTERY OF TESTS FOR MEASURING DRIVER SKILL
AND SAFETY AND PROVIDING AN INDEX OF IMPAIRMENT



direct and valid measure of impairment. Also, it would be ideal if experts were to agree that the behavioral tests were more closely related than BDC to Driver Skill and Safety--which is again more a matter of judgment than empirically verifiable fact.

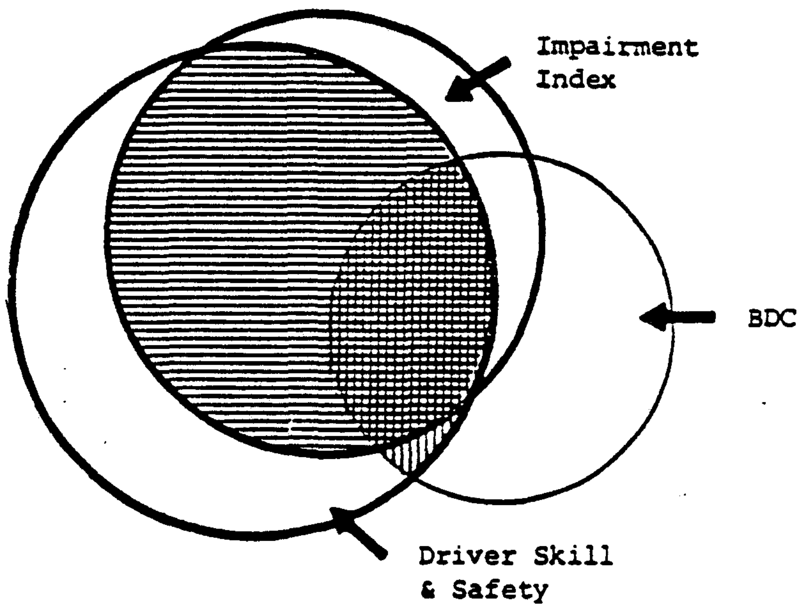
Properly constructed, a battery of distinct behavioral tests will yield an index of impairment that has the advantages of multiple operationalism (Webb, Campbell, Schwartz, and Sechrest, 1966). This means that the aggregate score, taken across tests, will provide a better measure of Driver Skill and Safety and impairment than any individual test. This situation is shown in Figure 7. This figure shows the happy case in which the index is valid. An important property of such an index would be that there would be relatively few sources of error to which experts could point, i.e., low index scores mean degradation in Driver Skill and Safety.

Of course, even a battery of behavioral tests will not capture all of what is meant by Driver Skill and Safety. Some aspects of skill and safety may not be measurable, either for practical or theoretical reasons, with behavioral tests. In developing a battery, therefore, it is critical to ensure that the most important aspects of skill and safety are captured in the index and that these include those factors most likely to be affected by drugs or other impairers.

Although the construction of such a behavioral index of impairment must be largely judgmental, it should be subject to empirical validation, a topic we will discuss again later. The important property of such an index is that it is intended as a measure of an endogenous condition of the driver: namely reduced capability in regard to Driver Skill and Safety. With this type of an index, we are trying directly to assess the internal capabilities and capacities of the driver, rather than to predict exogenous events. With such an index we want to be able to argue convincingly that the driver is impaired because his capabilities are reduced, and that that is what the index measures. It is a quite different matter to determine what the strength of the relationship is between reduced capabilities, BDC, accidents, test-course performance, and other items that may be of interest for one reason or another.

FIGURE 7

IMPAIRMENT INDEX AND DRIVER SKILL AND SAFETY



2.5 Validation of a Behavioral Index of Impairment

We have defined impairment as reduced capability, and it is primarily a matter of judgment to determine whether a reduction in a test score reflects a decrement in some underlying capability relevant to Driver Skill and Safety. Similarly, in most instances, only judgment can relate the underlying capability to accident risk. The validity ratings in this study reflect a consensus of the small group who participated in evaluating alternative behavioral tests.

Snapper, O'Connor, and Einhorn (1975) discuss in somewhat more detail how one validates an index of this type. The first step is verification of the "internal validity" of the index, which means that experts agree that the index reflects Driver Skill and Safety. A follow-on study would require that representatives of several different disciplines and enforcement areas be queried about validity of the index as a measure of impairment. An index would be carefully scrutinized to determine if the scoring procedure and component tests could be improved so that the testing procedure itself is as sound as possible.

The second type of validation is "external validity" of the index: this means essentially hypothesizing and then verifying empirical relationships involving the index. A behavioral index of impairment may be expected to have modest but reasonably consistent patterns of correlation with accidents, test-course performance, driver ratings, BDC, etc. The existence of weak but positive correlations should be expected, and the weakness thereof should not be construed as invalidity of the index for reasons we have discussed.

Most important, readers of this report who, on the basis of existing data or intuition, anticipate low correlations should not despair at the outset that this implies that no valid behavioral index can be developed. On the contrary, that finding would be welcome grist for our mill.

2.6 Criteria for Evaluating Indices of Impairment

We have already discussed one type of criterion: namely, the "ultimate criterion." The problem of the ultimate criterion was that of deciding what a behavioral index should be intended to measure. We argued that a behavioral index of impairment should be designed and intended to assess Driver Skill and Safety, even though that concept might be difficult to operationalize fully.

In this section, we consider a related but more concrete problem: namely, if one were to develop several alternative behavioral indices of impairment, which criteria would one use to decide that one index is better than another? This is a very practical problem, since it would be desirable to be able to deploy the battery at the roadside and/or in the station without requiring elaborate equipment and a cadre of trained psychologists.

Even if we were to assume that it would be possible to identify tests that could be used to measure validly Driver Skill and Safety, there might be no battery of tests that could satisfy certain practical criteria. Although this study stops short of actually generating and evaluating batteries, it does explicitly identify practical criteria for individual tests that might be incorporated into a battery.

Some of these criteria are validity criteria, which are used to assess whether tests are likely to measure what we want them to (i.e., Driver Skill and Safety). Other criteria pertain to reliability, ease of operational use, safety, and diagnostic value of the test.

In the remaining sections of this report, we will be considering various behavioral tests that are plausible candidates for inclusion in the type of index we have been describing. We will be explicitly evaluating these various behavioral tests using the criteria described above. The purpose of this evaluation procedure is to determine whether there are some testing approaches that might be further considered as part of a test battery, and whether it seems worthwhile to proceed further in developing one or more prototypical batteries.

3.0 CRITERIA, TESTS, AND EVALUATION

Attempts to identify useful behavioral tests by reviewing the literature have not yielded compelling results. Various individual tests, and batteries of tests, have been proposed as behavioral measures of impairment; but there is astonishingly little consensus of opinion either in the literature or among reviewers about how "good" they are. The literature, in fact, suggests that all behavioral testing procedures have problems, lending credence to the argument that good behavioral tests of impairment will never be developed.

The present study suggests a different interpretation and conclusion, however. It is clear that behavioral tests exhibit a variety of flaws. It is also clear that there are many different criteria one might want a behavioral test to satisfy, and that one will always be able to identify one or more criteria that a given test does not satisfy. This invites careless inferences about whether "good" behavioral tests are feasible. To illustrate, there are many different criteria for defining "validity," and a critic of a given test will always be able to identify <u>some</u> sense in which <u>any</u> given test is "invalid." Analogous reasoning for different types of criteria would suggest that no behavioral test would satisfy them either. But does it follow that "good" behavioral tests of impairment cannot be developed? We think not.

As we view the problem of assessing behavioral tests, any behavioral test (like any chemical test) is imperfect in one or more respects, and a critical technical problem is delineating multiple evaluative criteria. We wanted, for example, to specify multiple validity criteria in order to identify the particular criteria that a test <u>satisfied</u> (and those it violated), rather than grossly defining a test as "invalid." We wanted to identify specific validity criteria, specific reliability criteria, and other specific criteria with which to assess the goodness of behavioral tests overall.

We also wanted to quantify tradeoffs among criteria implied by different tests, so that the reader of this report could determine the comparative strengths (and weaknesses) of the tests. Using this approach, we hoped to determine whether there are behavioral tests good enough overall to show promise for future development.

3.1 Evaluation Approach and Methodology

The major question addressed by the evaluation was: Are there any behavioral tests that are "reasonably good" overall? Recognizing that there is no perfect behavioral test of impairment, the methodology (1) comparatively scored tests according to each of several criteria, and (2) then aggregated scores across criteria to arrive at an overall measure for each test.

The evaluation of these tests was necessarily subjective because of the paucity of data on the effects of many of the drugs, and even fewer data relating drug use to accident risk. All subjective judgments were made by the study team based on extensive reviews of available literature, discussions with other experts, and several years of research on this topic.

We used an explicit quantitative method for evaluating tests. A mathematical model, widely used in evaluation research, called Multiattribute Utility (MAU) theory, provided the methodology for assessing tests and for aggregating scores across criteria (Edwards, 1977; Keeney and Raiffa, 1976). MAU models had previously been used to evaluate chemical test procedures for identifying impaired drivers (Edwards, Goodman, and Snapper, 1971). MAU evaluation models can become extremely complex, especially if there are interactions among criteria in regard to defining an overall measure of goodness; we assumed an additive linear MAU model. In this form, the application of MAU theory to the evaluation of behavioral tests can be described in terms of the following steps:

Step 1. Identify all evaluative criteria. In this step, the various validity, reliability, and practical criteria (e.g., safety, ease of use) are identified. Criteria are defined so that they are conceptually mutually exclusive; that is, they measure different quality attributes of a test. The MAU evaluation model should include criteria assessing each major attribute of a test. Recognizing that all criteria are not equally important, we sought to be as comprehensive as possible in incorporating criteria in the MAU model.

Step 2. Identify a representative set of tests. Ideally, each conceivable behavioral test would be explicitly evaluated. Because of the many different types of tests, and the very large number of variations of most tests, this was impossible. The approach used was to identify a representative set of tests--not just those that appeared best or most promising at the outset of this study, or which had the widest popular support--under the assumption that our general results would generalize reasonably well to tests that might be identified in the future.

Step 3. Assess each test on each criterion. In this step, tests are comparatively assessed on a criterion-by-criterion basis. The assessments are made on scales standardized to the 0-100 interval. The 100-point represented the best plausible degree of criterion satisfaction, irrespective of whether any test could be identified that so fully satisfied the criterion to the minimal plausible extent. Definitions of "best plausible" and "worst plausible" were developed to ensure consistency in interpretation.

All tests were comparatively assessed and assigned scores between the 0- and 100-point extremes. The scores assigned to each test are judgmental, but linked to the best available data in the literature.

Step 4. Identify stakeholder groups and specify importance weights for evaluative criteria. Discussions with several interested experts indicated that there were likely to be systematic differences in terms of perceptions about the relative importance of different criteria. The police officer who would deploy a test might reasonably regard "ease of use" criteria as more important than a basic research might, for example.

Importance weights define the "significance" of different criteria. In MAU evaluation models, criteria with high weights are most salient in determining the overall score for a test. Two sets of importance weights were developed: one set represented our collective impression of research community perception, and the second set represented the perspective of the police officer or other test administrator. In each case, weights are normalized to sum to 1.0 (unity). Two sets of weights were used to determine to what extent, if any, differing perceptions about the relative importance of the criteria would influence overall results in regard to the evaluation of the tests.

Step 5. Compute overall MAU scores. The overall MAU score for a given test is readily computed, and is the weighted linear average of the scores on each criterion. That is, the score for a given test, on a given criterion, is multiplied by the importance weight for that criterion. The weighted scores for the various criteria are then summed to yield the overall score for the test, aggregated across criteria.

Step 6. Interpretation of MAU results. MAU evaluation results, as defined in Step 5, are overall scores aggregated across all evaluative criteria. Because weights are normalized to sum to 1.0, overall scores for each test will fall between 0 and 100. Generally, a test that receives a fairly high score must reasonably satisfy at least the most important criteria; conversely, tests receiving fairly low scores are generally those that fail to satisfy at least some of the more important criteria.

In interpreting results it is useful to consider disaggregated data, i.e., data for each test on a criterion-by-criterion basis. The MAU evaluation model, of course, permits display of disaggregated results; these are shown in tables in this report. Careful inspection of those tables will show just why a given test scored high (or low) overall. These tables also indicate the comparative strengths and weaknesses of different tests.

Disaggregated results are shown in part for heuristic purposes. We selected specific test configurations without attempting to "fine-tune" them or reconstitute them in order to make them "look good" in this evaluation. Yet it is clear that many of these tests may be improved upon in a more extended research effort. Disaggregated results are shown to enable the reader to determine the extent to which certain tests might be improved in relation to criteria they failed to adequately satisfy. That is, the results of the MAU evaluation might suggest directions in which possible behavioral tests might be improved in the future.

3.2 Development of Evaluative Criteria

Although the primary purpose of the present study was to identify "practical" criteria for behavioral tests of impairment, it became clear early in the study that any test would have to satisfy some distinctly different audiences. These audiences included the research community, enforcement officials, and the adjudicatory and legal community. The existence of these different audiences--and their different viewpoints--suggested that a variety of legal and technical as well as wholly practical criteria would have to be considered in assessing behavioral tests.

In identifying criteria it was clear that the different audiences differed primarily in their emphasis on different criteria. That is, we anticipated developing a list of criteria that everyone would agree were relevant, although they might disagree about their relative importance. Three very broadly defined categories of criteria were identified: technical, practical, and legal.

Technical criteria included validity and reliability criteria, and attempted to represent the issues of most interest from a research perspective. Different relevant senses of "validity" and "reliability" are reflected in the criteria. The second broad category consisted of practical criteria that a test--irrespective of its technical merits--would have to satisfy reasonably before it would be developed and deployed. Generally, these criteria referred to the use of the test in the field, including suitability of instrumentation and acceptability to the subject and the officer.

The legal criteria we identified differed from technical and practical criteria in one critical respect: they represent conditions or procedures that any behavioral test must satisfy if it is to be used successfully. However, these criteria pertained primarily to how tests would actually be used in the field and in the courtroom rather than which tests should be used. Moreover, the legal criteria we identified did not preclude deployment of behavioral tests and tended to be indicative of approaches by which such tests could be successfully used. Thus, the legal criteria were in general not crucial to the initial selection and evaluation of behavioral testing procedures considered in this report, and were not explicitly used to evaluate tests in this study. However, because the legal criteria would be relevant to further development and deployment of tests subsequent to this study, they are discussed in an appendix to this report.

The specific criteria shown and defined in Table 1 were identified through relevant literature (e.g., Burns and Moscowitz, 1977), discussions with expert consultants, and discussions with NHTSA staff. These criteria are not defined in an absolute sense; a test to be useful does not need to achieve each of these. Rather, they define attributes along which behavioral tests can be better or worse.

Several of the validity criteria relate closely to the discussion in Section 2.0 For example, the criteria indicate that a behavioral test should be related to driving skill (1.1.1) and to accidents (1.1.2). Other criteria suggested in Section 2.0 are more applicable to batteries of test rather than individual tests, e.g., tests should cover the range of driving skills and behaviors. The validity criteria also include ability to detect impairment produced by multiple causes. For example, a test is better if it detects impairment produced by tranquilizers as well as that produced by alcohol. Other validity criteria relate to the test's capability to discriminate impaired drivers from the general population of drivers without an individual baseline measurement, and to knowledge about what the test does not measure.

The reliability criteria include standard test-retest reliability as well as reliability criteria directly related to the implementation of behavioral tests. These latter criteria include lack of effects produced by changes in testing environment, by subject motivation and practice, and by test equipment. An additional important reliability criteria relates to reliability across demographic populations.

TABLE 1

CRITERIA FOR EVALUATING
BEHAVIORAL TESTS OF IMPAIRMENT

1.0 Technical

1.1 Validity

1.1.1 Related to driving skills and competence -- Extent to which a test relates to the multiple functional abilities of "normal" competent drivers. This includes the extent to which an ability measured by the test is important in driving and the range of driving-related abilities reflected in the test.

1.1.2 Predictive validity (or potential) with respect to accidents -- Extent to which performance on test predicts probability of accident. This also includes the potential of the test with respect to predictive validity based on expert opinion, current literature, and face validity with respect to the relationship between the test and impaired behavior.

1.1.3 Identifiability of impairing conditions not measured by test -- Extent to which it is possible to determine and document which forms of impairment are not measured by test. This information is necessary for the officer to make a decision about whether an impairing condition (such as fatigue) existed but was not reflected in the roadside administration of the test. This information is also needed by courts, in order to determine when negative behavioral test results should be set aside, rather than be used as evidence that counters the officer's description of behind-the-wheel performance.

1.1.4 Single-test interpretation -- Extent to which the officer and/or courts can interpret results based upon single administration of the test. This includes especially the extent to which the test yields interpretable results without having established a baseline for each individual. (For instance, a test could be interpretable on a single administration if population norms were well-established, especially if it were "obvious" that any normal individual should be able to pass the test.)

1.1.5 Detects impairment from multiple causes -- Extent to which test can validly assess impairment from any cause.

1.2 Reliability

1.2.1 Test-retest reliability -- The correlation (or consistency) between test scores on initial and subsequent administration; the inherent degree of replicability.

1.2.2 Reliability across ambient environmental/physical conditions -- Including the applicability of the test under inclement weather conditions, independence of results with respect to time of day or season.

1.2.3 Non-alterable (stable) performance -- The extent to which the test yields "true" results that are not affected by a variety of factors, including motivation, intent to perform poorly, practice, or passive effects of time.

1.2.4 Reliability of equipment -- Reliability of any required equipment and lack of interference from ambient conditions. Lack of equipment sensitivity to operator error, misadjustment.

1.2.5 Reliability across demographic populations -- Extent to which the test is culture-free and comparable results can be obtained across different demographic populations, which are similar in regard to the underlying skills/behaviors measured by the tests. For instance, it should not be necessary to "recalibrate" the test for different socio-economic groups. (Note: this overlaps partially with validity criteria, but is defined here as a reliability factor distinct from validity.)

2.0 Practical

2.1 Ease of Operational Use

2.1.1 Duration of test procedure -- The length of time it requires the officer to administer the test. This criterion is especially important in regard to tests administered roadside (as opposed to in the station). Duration of testing could refer to a quick screening battery, if a sequential test procedure is possible.

2.1.2 Minimal equipment -- Apart from problems of reliability and cost, the amount of equipment required should be minimized. This criterion is especially relevant in regard to roadside tests, where instrumentation is likely to be cumbersome to carry and inconvenient to use.

2.1.3 Ease of scoring -- Extent to which test is self-scoring, requiring no judgment or computation by the administrator; easily interpretable by the administrator (i.e., the impaired versus unimpaired decision can be made); and provides continuous scores, which can be calibrated for the impaired/unimpaired decision.

2.1.4 Minimal training for test administrators -- Amount of training required for officer validly and properly to administer tests and to be able to testify in court.

2.1.5 Degree of cooperation required of subject -- Extent to which subject must be willing and able to follow instructions.

2.2 Safety

2.2.1 Safety to suspect -- Degree to which the test procedure does not pose any potential physical or mental risk to the subject; test should be as "natural" as possible.

2.2.2 Safety for law enforcement officer -- Degree to which administering the test does not pose a possible threat to the officer. This includes the test not requiring prolonged, concentrated attention by the officer on test administration rather than the subject, not limiting access to weapons, and not antagonizing passengers or bystanders because it is too long, too comical, or too conspicuous.

2.3 Diagnostic value of the test -- Extent to which the test can distinguish among different causes or sources of impairment. (For example, whether the test or test battery produces a "signature" characteristic of a particular class of drugs, or whether it distinguishes between alcohol-only impairment versus impairment caused by a drug but the subject has taken comparatively small amounts of alcohol to mislead the officer.)

The practical criteria focus on the operational use of the test including duration, scoring, equipment, training, and subject cooperation. They also include the safety of administering the test, both to the subject and to the administrator; and the ability of the test to differentiate among causes of impairment.

3.3 Identification of Behavioral Tests

Use of behavioral tests to measure impairment has a long history of research during which numerous tests have been tried. Jellinek and McFarland (1940) provide an early review of this research, most of which has focused on alcohol-caused impairment. More recent reviews, including Perrine (1974), Austin, Sterling-Smith, Macari, and Lettieri (1977), Willette (1977) and the NHTSA-supported bibliography (Joscelyn and Donelson 1979; Joscelyn and Maikel, 1977; Veldkamp, Donelson, and Joscelyn, 1980a, 1980b), suggest the expanded focus on other drugs in addition to alcohol. These reviews, and the literature cited therein, served as our point of departure for identifying potential behavioral tests.

We also extended the scope of our search for behavioral tests to identify potential tests that had not necessarily been used previously specifically to test driver impairment. In this extended search, conducted using the DIALOG system, we identified various types of human performance, and performance measurement tests without any requirement for use to test impairment. From this search, we identified a large number of behavior tests covering sensory, perceptual, cognitive, and motor processes for further evaluation.

These tests, categorized by the type of behavior required, are listed and described in Table 2. Also given are the numbers of the references in which these tests have been used.

The literature search and review indicated that the effects of a large number of drugs on behavior have been investigated. Most drugs, however, have had very few investigations on a minimal number of behaviors. Table 3 provides references for tests of the effects of various drugs.

TABLE 2

BEHAVIORAL TESTS AND RELATED REFERENCES

TEST	TYPE OF BEHAVIOR	DESCRIPTION	REFERENCES
Auditory Frequency Differential Threshold	Sensory-Perceptual	The ability to state which of two pitches is higher in pitch or if the two pitches differ. The threshold is the minimal difference perceivable. The S is merely asked to make the judgment. A staircase method (ascending and descending series) is usually employed.	31, 35, 93, 172
Auditory Intensity Differential Threshold	Sensory-Perceptual	The ability to state which of two sounds is louder. The threshold is the minimal difference at which such discrimination is possible. The S is merely asked to make the discrimination. Typically presented in 5 ascending and 5 descending 1-db step intervals.	31, 93, 172
Auditory Threshold	Sensory-Perceptual	The minimal sound energy that is perceivable. Measured in decibels. The S states whether he/she can detect a sound or not. Usually presented in 5 ascending and 5 descending 2-db step intervals.	31, 93, 172
Autokinesis	Sensory-Perceptual	The perception of apparent movement of a stationary point source of light viewed in the dark. Subjects are asked to identify the square in a grid into which the light appears to fall (or some background).	153, 188

TABLE 2 (Continued)

BEHAVIORAL TESTS AND RELATED REFERENCES

TEST	TYPE OF BEHAVIOR	DESCRIPTION	REFERENCES
Brightness Sensitivity	Sensory-Perceptual	This test determines S's ability to discriminate between lights of varying brightness. S is shown two lights and must indicate which is brighter. Performance is measured by the difference in brightness at which S can reliably discriminate a difference.	15, 91, 93, 203
Color Discrimination	Sensory-Perceptual	The faculty by which colors are perceived and distinguished. Characterized by the attributes of hue, saturation, and brightness. Measured by (1) minimum amount of a color added to a white light to match another color, (2) sensitivity of one color relative to another, and (3) naming of colors.	2, 20, 39, 44, 170
Critical Flicker Fusion	Sensory-Perceptual	The transition point at which an intermittent, rapidly flickering light source is first perceived as being continuous or fused. Measured in flashes/sec, or cycles/sec.	11, 60, 69, 78, 79, 89, 104, 108, 110, 115, 118, 123, 134, 135, 147, 156, 163, 164, 178, 184, 186, 198, 202
Delayed Auditory Feedback	Sensory-Perceptual	S is given a series of reading, counting, and arithmetic tests requiring verbal responses which are played back through earphones at slightly higher intensity and with a delay. Performance is measured by the average number of correct responses per error. Usually, only two minutes of each recorded test is graded.	43, 53, 55, 56, 82, 83, 84, 99, 106, 132

TABLE 2 (Continued)

BEHAVIORAL TESTS AND RELATED REFERENCES

TEST	TYPE OF BEHAVIOR	DESCRIPTION	REFERENCES
Duration of Archimedes Spiral Afterimage	Sensory-Perceptual	The perception of rotation in the opposite direction after exposure to a rotating stimulus. The rotating stimulus is an Archimedes Spiral; the measure is the length of such perception.	35, 104, 122
Dynamic Visual Acuity	Sensory-Perceptual	This test measures S's ability to discriminate moving targets. It can use letters projected on a screen or Landolt rings. Performance is measured by the finest discrimination S can make.	16, 20, 24, 38, 73
Glare Recovery	Sensory-Perceptual	The ability to make visual distinctions in a surrounding field following the presentation of any intense light. Measured by (1) elevation in minimum intensity required to perceive a white shape following a bright light, or (2) time required to perceive direction of a dimly lighted arrow following bright light.	3, 20, 159, 217
Nystagmus	Sensory-Perceptual	With one eye covered, S follows movement of light or object with other eye without moving head. Performance is measured by jerking movement of eye.	4, 13, 26, 27, 59, 63, 75, 80, 81, 134, 135, 157, 163, 167
Peripheral Vision	Sensory-Perceptual	Similar to visual field, except that the task is one of signal detection. Measured by false alarms/hits of horizontal field signals.	115

TABLE 2 (Continued)

BEHAVIORAL TESTS AND RELATED REFERENCES

TEST	TYPE OF BEHAVIOR	DESCRIPTION	REFERENCES
Static Visual Acuity	Sensory-Perceptual	This test measures S's ability to distinguish relatively small spatial intervals between nonmoving objects. It can use a Snellen Chart, Landolt ring, or the Bausch and Lomb Ortho-Rater. Performance is measured by the smallest interval which S can reliably discriminate.	1, 20, 88, 103, 138, 141, 200
Two-Point Tactile Discrimination	Sensory-Perceptual	S is touched lightly with two pointed stimuli. The distance between the stimuli is varied and S must indicate whether he/she feels one or two stimuli. Performance is measured by the minimum distance at which S indicates two stimuli are felt.	27, 93
Visual Field	Sensory-Perceptual	The extent of the lateral visual field while the eyes are fixated on a point straight ahead. This is measured directly as the absolute horizontal length at which objects can be perceived.	14, 20, 39, 93, 105, 159, 170
Color Naming	Perceptual-Cognitive	Numbers 10-59, in random order, in four colors are on a card in rows presented to S. S must find a sequence of ten numbers, beginning with a specified number, and verbally report the color of each number as quickly as possible. Performance is measured by the time required to perform the task and by the number of errors.	27

TABLE 2 (Continued)

BEHAVIORAL TESTS AND RELATED REFERENCES

TEST	TYPE OF BEHAVIOR	DESCRIPTION	REFERENCES
Distance Estimation	Perceptual-Cognitive	Conducted either from within an actual vehicle, or from a simulator, S is asked to estimate (1) the gap between his/her car and car ahead, (2) distance from a curb, (3) the gap between two posts, or (4) distance between front tire and white line.	37, 107, 177, 211
Time Estimation	Perceptual-Cognitive	This test can be performed in several ways. S can be given a certain time and be asked to respond when that time passes. Alternatively, S can be given two signals at a certain time interval and be asked to estimate the time between the signals. Performance is measured by comparing S's responses with actual time elapsed.	57, 69, 72, 107, 108, 119, 177, 181, 195, 210, 215, 216
Arithmetic	Cognitive	Performance is usually measured by the number of additions completed in a brief test, though it is not necessary to use addition.	17, 29, 30, 34, 41, 47, 85, 120, 128, 195, 208, 214, 215, 216, 221
Code Substitution	Cognitive	Typically employs the Digit Symbol Substitution Subtest of the WAIS. The S is asked to substitute arbitrary symbols for letters as rapidly as possible in a brief time span, such as two minutes. Number correct is measured.	202

TABLE 2 (Continued)

BEHAVIORAL TESTS AND RELATED REFERENCES

TEST	TYPE OF BEHAVIOR	DESCRIPTION	REFERENCES
Complex Counting	Cognitive	S is instructed to count by n, where n is some integer, e.g., "Count by Sevens." S is not allowed to perform additions out loud. Occasionally, a certain pattern of subtraction is included. The counting is usually fast, with errors and latency being measured.	18
Counting Backwards	Cognitive	S must count backward beginning at some specified number. A common variation of this is to count backwards by threes or even sevens. Performance is measured by speed and accuracy with which task is performed.	10, 27, 87, 158, 167, 221
Digit Memory	Cognitive	Adapted from the digit span test of the Wechsler-Bellvue Intelligence Test. S is presented a series of digits, which then must be recalled in backwards order. The number correct in backwards order is recorded.	66, 121, 138, 195, 221
Serial Performance	Cognitive	This test uses a device with five toggle switches in a row and a light. All switches are initially in the center position. S must move the switches to up and down positions until when they are in the correct up-down sequence, the light will come on. Performance is measured by the time required to reach the correct switch sequence.	27, 165, 204

TABLE 2 (Continued)

BEHAVIORAL TESTS AND RELATED REFERENCES

TEST	TYPE OF BEHAVIOR	DESCRIPTION	REFERENCES
Sorting	Cognitive	S is required to place wooden tablets into holes on the basis of shape and/or color.	69, 147
Stroup Test	Cognitive	The S is quickly presented a series of cards. On each card is the name of a color written in a different color ink (e.g., "brown" printed in green). The task is to give the color of the word, and not the word itself. The total number correct is recorded.	32, 62, 73, 74, 175, 195
Tongue Twisters	Cognitive	S must repeat words which are difficult to say distinctly, e.g., "Methodist," Episcopalian," "sophisticated statistics." Performance is measured subjectively by S's ability to say these words clearly and distinctly.	27
Choice Reaction Time	Perceptual-Motor	This task can be implemented in a number of ways. Two or more stimuli can be presented to S, each of which has a separate, appropriate response. When a stimulus is presented, S must decide which stimulus it is and respond appropriately as quickly as possible. Performance is measured by both response errors and response time.	36, 58, 60, 67, 78, 111, 113, 114, 115, 116, 117, 123, 124, 125, 127, 129, 142, 143, 162, 163, 164, 180, 185, 186, 219, 220

TABLE 2 (Continued)

BEHAVIORAL TESTS AND RELATED REFERENCES

TEST	TYPE OF BEHAVIOR	DESCRIPTION	REFERENCES
Compensatory Tracking	Perceptual-Motor	This task, typically implemented on a CRT requires S to maintain a visual signal on a target. The signal receives random input taking it off target and S, controlling the velocity, attempts to keep the signal on target. Performance is measured by some error function, e.g., average squared error, indicating how long and how far the signal was off target.	40, 51, 104, 127, 132, 133, 179
Critical Tracking	Perceptual-Motor	This task, typically implemented on a CRT, requires S to maintain a visual signal on a target. This signal moves away from the target with a force that is an increasing function of the distance of the signal from the target and of time. S controls the displacement of the signal to keep on target (usually a region). Performance is measured by the level of task difficulty at which S fails to keep the signal on the designated target.	51, 187
Pursuit Tracking	Perceptual-Motor	This task, typically employing a pursuit rotor, requires the S to keep a loose-handled stylus in contact with a small disc rotating clockwise at various speeds (usually 50 rpm). The time of contact is recorded, typically over several trials of short (10 secs) duration. A second kind of task is the pursuit meter. Here, the subject is asked to use a steering wheel or other	25, 35, 43, 68, 82, 85, 99, 106, 110, 112, 113, 114, 130, 132, 142, 143, 162, 183, 210

TABLE 2 (Continued)

BEHAVIORAL TESTS AND RELATED REFERENCES

TEST	TYPE OF BEHAVIOR	DESCRIPTION	REFERENCES
		device to track oscilloscope traces of varying complexity. A digital error score between the instrumental problem and motor response is measured.	
Simple Reaction Time	Perceptual-Motor	This task can be implemented in a number of ways. S must respond as quickly as possible to a prespecified stimulus, usually either auditory or visual. Performance is measured by the time between presentation of the stimulus and the response.	11, 22, 23, 29, 34, 58, 60, 61, 67, 71, 110, 116, 126, 155, 156, 195, 199, 200, 219
Tracing	Perceptual-Motor	S must trace with pencil a path through a maze on paper. Performance is measured by the time to complete the maze and the number of wrong turns in the maze.	27
Bender Drawing Test	Perceptual-Motor	This consists of nine standard simple geometrical figures presented on cards. The subject is merely instructed to copy the figures. S can use as much time as needed, and can view the figure during the entire copying time.	46, 167
Digit-Symbol Substitution Task	Perceptual-Motor	Pairs of one letter and one digit are presented to S, one at a time, at a fixed rate. Then the letters are presented in a row. S must respond with the corresponding numbers in the correct order as quickly as possible. Performance can be measured as percent correct responses within a fixed time	10, 18, 29, 41, 61, 71, 77, 92, 109, 120, 130, 138, 139, 145, 166, 210, 215, 216, 219, 220, 221

TABLE 2 (Continued)

BEHAVIORAL TESTS AND RELATED REFERENCES

TEST	TYPE OF BEHAVIOR	DESCRIPTION	REFERENCES
		period, response time, or a combination of correct responses and response time.	
Letter Cancellation (Bourdon)	Cognitive-Motor	S is given written text and must mark one or more specified letters as quickly as possible for a fixed time. Performance is evaluated by the number or percentage of letters correctly marked, the total amount of text covered, and/or the percentage of errors.	27, 120, 138, 158, 195, 196, 212, 220, 221
Minnesota Spatial Relations Test	Cognitive-Motor	This task uses boards with 58 cutouts of various sizes and shapes. S must move the cutouts from one board to another on which each cutout is differently located. Performance is measured by the time required to complete the task.	224
Trail Making Test	Cognitive-Motor	This is a timed, paper and pencil test with two parts requiring approximately five to seven minutes to administer. S must locate and connect symbols on a page. The first part requires S to locate and connect in the proper sequence circles numbered 1 to 25. This part demands visually coordinated rapid motor movements. In the second part, S connects circles in ascending sequence alternating between numbers and letters. S works until correct performance is achieved and is scored on the basis of time.	7, 176, 214

TABLE 2 (Continued)

BEHAVIORAL TESTS AND RELATED REFERENCES

TEST	TYPE OF BEHAVIOR	DESCRIPTION	REFERENCES
Finger Count	Motor	S must touch and count each finger successively with thumb and then reverse. "One" is thumb alone, "two" is index finger, etc. Thus, the sequence goes 1-2-3-4-5-5-4-3-2-1. Performance is measured by judgment of S's execution of task.	27
Finger-to-Finger	Motor	In this test, S stands with eyes closed and arms spread. The index fingers are then brought together in front of S's face. Performance is measured by the accuracy with which the fingers are touched.	128, 167, 171
Finger-to-Nose	Motor	This task requires S to stand with eyes closed, and arms extended horizontally. S must then touch his/her nose with index finger alternating left and right hands. Performance is measured by ability to easily touch nose with fingers.	27
Grip Strength	Motor	Using a dynamometer, S must squeeze as hard as possible. Performance is measured by the force exerted.	5, 27, 45, 52, 76, 93, 204
Modified Romberg Test	Motor	In this test, S stands with one foot in front of the other, head tipped back, eyes closed, and hands at side. Performance is measured by the amount of body sway.	59, 76, 171

TABLE 2 (Continued)

BEHAVIORAL TESTS AND RELATED REFERENCES

TEST	TYPE OF BEHAVIOR	DESCRIPTION	REFERENCES
One-Hole Test	Motor	S grasps one prepositioned pin, moves it seven inches in a direction 45° away from the body, positions it into a hole with close tolerances and reaches to grasp the next pin. Score is the number of pins inserted per one minute trial.	182, 182a, 182b
One-Leg Stand	Motor	S stands with one leg held straight in front slightly elevated from the ground. Performance is measured by S's ability to maintain this position.	26, 27, 160
Picking up Objects	Motor	Small objects are placed on the ground near S. S must pick them up, one at a time. Performance is assessed subjectively.	27, 167
Romberg Test	Motor	In this test, S stands with feet together, head tipped back, eyes closed, and hands at side. Performance is measured by the amount of body sway.	12, 27, 59, 90, 167, 173
Tapping	Motor	S must tap a telegraph key as rapidly as possible. Performance is measured by the number of taps in a given time period.	18, 27, 29, 30, 41, 47, 66, 71, 77, 86, 89, 104, 111, 112, 113, 117, 143, 144, 155, 162, 166, 195, 199, 204, 219
Walk and Turn	Motor	S is required to walk a straight line touching heel to toe, turn and return in same manner. Performance is assessed subjectively.	26, 27

TABLE 2 (Continued)

BEHAVIORAL TESTS AND RELATED REFERENCES

TEST	TYPE OF BEHAVIOR	DESCRIPTION	REFERENCES
Wobble Board	Motor	A device used to measure balance-seeking behavior. The S stands on a platform mounted on a dual axis torsion rod system. The S is asked to gaze straight ahead while maintaining balance. The WB is attached to an electronic counter which measures amount of angular movement over a short period of time (20 secs).	43. 58, 68, 106
Divided Attention		Any perceptual situation in which two tasks must be performed simultaneously. The two tasks may be intra- or inter-sensory channel. Measured in a variety of ways; typically, the secondary task is some kind of tracking.	33, 48a, 51, 131, 146, 149, 150, 151, 152, 153, 154, 163, 174, 205
Field Dependence		The extent to which a person is capable of overcoming an embedding context in order to perceive relevant targets. Measured by time taken to scan a complex figure in which a target is embedded. Two standard measures are the Embedded Figure Test and the Rod and Frame Test.	8, 9, 19, 42, 65, 70, 140, 161, 191, 206
Vigilance		Any task, either auditory or visual, in which the subject is asked to respond to a lengthy presentation of stimuli. The stimuli occur randomly, and are either pure tones or blotches of light. RT following each stimulis is measured. The task is usually one hour long, or greater.	28, 29, 30, 67, 71, 77, 109, 126, 149, 156, 166, 189, 215, 216

TABLE 3

REFERENCES FOR TESTS OF EFFECTS
OF VARIOUS TYPES OF DRUGS

DRUG	REFERENCES
Alcohol	1, 2, 3, 4, 5, 7, 10, 12, 13, 15, 24, 25, 34, 37, 38, 39, 40, 52, 55, 56, 58, 59, 63, 68, 69, 72, 76, 81, 82, 83, 84, 86, 87, 89, 91, 93, 105, 111, 112, 113, 114, 116, 119, 120, 122, 124, 125, 126, 127, 128, 129, 130, 131, 132, 135, 142, 143, 147, 149, 152, 153, 155, 157, 158, 159, 160, 162, 163, 164, 170, 171, 172, 173, 177, 180, 185, 196, 203, 205, 212, 214, 220, 221
Antidepressants	29
Amitriptyline	114, 162
Desipramine	113
Imipramine	62, 74, 113, 215
Lithium	45, 129, 156
Nomifensine	216
Nortriptyline	83
Viloxazine	11, 74
Cannabis & Related Agents	
Marijuana	1, 2, 3, 24, 28, 31, 33, 35, 40, 43, 53, 104, 107, 131, 132, 133, 139, 144, 150, 151, 153, 177, 179, 182, 184, 187, 188, 189, 210
Delta-9-Tetrahydrocannibinol and other Tetrahydrocannabinols	34, 106
Hallucinogens	
LSD	75
Mescaline	75
Nitrous Oxide	61
Psilocybin	75

TABLE 3 (Continued)

REFERENCES FOR TESTS OF EFFECTS
OF VARIOUS TYPES OF DRUGS

DRUG	REFERENCES
Sedatives & Hypnotic Agents	
Barbiturates	47, 66, 146
Amobarbital	32, 71, 89
Pentobarbital	18, 175
Pentobarbitone Sodium	22
Phenobarbitone	155
Secobarbital	43, 80, 109, 178
Non-Barbiturates	
Benzoctamine	111
Benzodiazepines	22, 216
Diphenhydramine	10
Flurozepam	204
Fosazepam	204
Glutethimide	41, 155
Nitrazepam	204
Stimulants	18
Amphetamine	47, 214
1-Benzylpiperazine	30
Caffeine	158, 160
Dextroamphetamine	13, 30, 160
Methamphetamine	32, 175
Tranquilizers	17, 38, 79, 84, 88
Major	
Chlorpromazine	109, 123, 142, 186, 220
Haloperidol	37, 143, 208
Prochlorperazine	103
Sulpiride	186
Thioridazine	142, 186
Trifluoperazine	37
Hypnotics	99, 166

TABLE 3 (Continued)

REFERENCES FOR TESTS OF EFFECTS OF VARIOUS TYPES OF DRUGS

DRUG	REFERENCES
Tranquilizers (continued)	
Minor	
Bromazepam	186
Chlordesmethyldiazepam	219
Chlordiazepoxide	23, 37, 79, 85, 89, 141, 183
Clobazam	78
Diazepam	23, 69, 71, 80, 85, 92, 108, 120, 123, 127, 128, 129, 134, 135, 143, 147, 154, 199, 211
Lorazepam	14
Medazepam	116
Meprobamate	56, 88, 103, 138, 200, 221
Other Central Nervous System Agents	
Fenmetozole	68
Methadone	67
Valproate Sodium	195
Antihistamines	82, 112, 178
Clemastine	44
Dexchlorpheniramine	58
Tertenadine	110
Other Drugs	
Atenolol	16, 115
Atropine	124
Carbon Monoxide	174, 217
Drug Metabolites	36, 163, 164
Ethylbenzene	60
Gamma-hydroxybutyric Acid	134
Isoniazid	125
Mercury	117
Promethazine	118
Propranolol	115, 199
Trithiozine	135
Xylene	60

3.4 Evaluation of Behavioral Tests

Evaluation of the behavioral tests was quantified using the MAU approach described in Section 3.1. This quantitative evaluation, supplemented by the rationales for it presented below, required two types of inputs. The first is the importance weights for each criterion which define how much that criterion contributes to the overall potential of the test. The second is comparative assessments of the behavioral tests on each criterion. Both of these inputs were provided by the study team, after extensive review of related literature and discussions with consultants and NHTSA staff. These inputs are described in detail below.

3.4.1 Importance weights for criteria. Weights that are assigned to criteria must reflect the range of possibilities on the criteria. For example, if the range on criterion 2.1.1, duration of test procedure, is two minutes, the weight should be lower than if it is ten minutes. In general, weights should reflect how desirable it is to go from the worst possible on the criterion to the best possible. Thus, the first step in determining weights was to define the range on each criterion. In these definitions, we took into account both the range over which behavioral tests might vary, and a desirable range which might not be achieved by any existing test. Table 4 defines the range for each criterion. Note that the range for any criterion under which there are subcriteria is defined by the range from the worst on all subcriteria to the best on all subcriteria.

Two perspectives were adopted in assigning weights: that of law enforcement officers who would administer the tests, and that of researchers who would develop and validate the tests. The researchers perspective was meant to represent the viewpoint focusing on development of the most valid tests possible with relatively less emphasis on other criteria. Test administrators, on the other hand, would be concerned specifically with the practical matters of actually using the tests. The weights used for these perspectives are shown in Table 5. These weights are not meant to be definitive, nor do they suggest broad agreement among either researchers or test administrators.

TABLE 4

RANGE OF POSSIBILITIES FOR EACH CRITERIA

1.0 Technical Criteria

1.1 Validity

1.1.1 Related to Driving Skills and Competence

Worst - Entirely unrelated.
Best - Measures several functional abilities, each of which is an integral driving skill.

1.1.2 Predictive Validity with Respect to Accidents

Worst - Unpredictive.
Best - Existing evidence (including expert opinion) indicates the test is highly predictive of accidents.

1.1.3 Identifiability of Impairing Conditions Not Measured by Test

Worst - Impairing conditions not measured are undetermined.
Best - All impairing conditions not measured are known or can be determined.

1.1.4 Single-test Interpretation

Worst - Test requires complete set of individual baseline data.
Best - Test discriminates impaired driver from unimpaired population without individual baseline data.

1.1.5 Detects Impairment from Multiple Causes

Worst - Specific to single type of impairing agent.
Best - Detects impairment from any cause.

1.2 Reliability

1.2.1 Test-retest Reliability

Worst - No correlation between test and retest.
Best - Correlation is 1.0.

1.2.2 Reliability across Ambient Environmental/physical Conditions

Worst - There are conditions under which test cannot be administered, or if it is, results are invalid.
Best - Completely unaffected by conditions.

TABLE 4 (Continued)

1.2.3 Non-alterable Performance

Worst - Performance is easily altered because of practice, motivation, or other factors.
Best - Performance cannot be altered.

1.2.4 Reliability of Equipment

Worst - Equipment failure is quite likely, and difficult to detect and repair.
Best - No chance of equipment failure.

1.2.5 Reliability Across Demographic Populations

Worst - Performance must be calibrated for multiple demographic groups.
Best - No relationship between test performance and demographic variables.

2.0 Practical Criteria

2.1 Ease of Operational Use

2.1.1 Duration of Test Procedure

Worst - Five minutes.
Best - No time.

2.1.2 Minimal Equipment

Worst - Equivalent in size to a large piece of luggage.
Best - None required.

2.1.3 Ease of Scoring

Worst - Entirely judgmental with little explicit guidance.
Best - Self-scoring requiring no judgment or interpretation.

2.1.4 Minimal Training for Test Administrators

Worst - Requires testing of administrator and certification with extensive formal training.
Best - None required.

2.1.5 Degree of Cooperation Required of Subject

Worst - Requires extensive voluntary physical actions that may be uncomfortable, unpleasant, or embarrassing.
Best - Requires no voluntary cooperation or overt actions.

TABLE 4 (Continued)

2.2 Safety

2.2.1 Safety to Suspect

Worst - Poses some physical and/or mental risk to subject.
Best - Poses no risk to subject.

2.2.2 Safety for Law Enforcement Officer

Worst - Requires prolonged, concentrated attention by officer, limits access to weapons, and antagonizes bystanders.
Best - No threat to officer.

2.3 Diagnostic Value of Test

Worst - Does not discriminate among any causes of impairment.
Best - Discriminates between at least two sources of impairment.

TABLE 5

CRITERIA WEIGHTS FROM PERSPECTIVES OF RESEARCHERS AND TEST ADMINISTRATORS

CRITERIA	NORMALIZED WEIGHTS	
	RESEARCHERS	ADMINISTR'S
1.0 Technical	.69	.34
1.1 Validity	.55	.44
1.1.1 Driving Skills	.42	.11
1.1.2 Predicting Accidents	.11	.18
1.1.3 Identifying Conditions Not Measured	.06	.07
1.1.4 Single-test Interpretation	.28	.43
1.1.5 Multiple Causes	.14	.21
1.2 Reliability	.45	.56
1.2.1 Test-retest	.07	.04
1.2.2 Ambient Conditions	.04	.26
1.2.3 Non-alterable Performance	.37	.43
1.2.4 Equipment	.02	.22
1.2.5 Demographic Populations	.49	.04
2.0 Practical	.31	.66
2.1 Ease of Use	.68	.58
2.1.1 Duration	.40	.51
2.1.2 Equipment	.04	.05
2.1.3 Scoring	.10	.13
2.1.4 Training	.05	.06
2.1.5 Cooperation of Subject	.40	.25
2.2 Safety	.23	.38
2.2.1 Suspect	.33	.17
2.2.2 Officer	.67	.83
2.3 Diagnostic Value	.09	.04

Note: Weights at each level are normalized to sum to one. Weights may not sum to exactly one because of roundoff.

The researchers put relatively high weight on test validity and reliability. Of the validity subcriteria, relevance to driving skill (1.1.1) and single-test interpretation (1.1.4) received the highest weights. Predictive validity (1.1.3) received a relatively low weight, not because it is not important in an absolute sense, but because as shown by the arguments in Section 2.0, it has a relatively small range. That is, it does not effectively differentiate among tests because differences among tests are small or unknown. High predictive validity is not expected from any single behavioral test. Much of the weight on reliability is because of non-alterable performance (1.2.3) and reliability across demographic populations (1.2.5).

The weights assigned from the perspective of test administrators show somewhat different considerations. Here the practical criteria, ease of operational use, and safety become relatively more important. From this perspective there is less concern for validity, particularly relevance to driving skills (1.1.1). Reliability is also less important, with reliability across demographic populations (1.2.5) being relatively less important, and reliability of equipment (1.2.4) and across ambient conditions (1.2.2) being relatively more important.

3.4.2 Comparative assessments of tests. Quantitative scores were assigned to each test on each criterion. These assessments were made on a 0 to 100 scale where 0 is defined as the worst feasible level of the criterion and 100 is the best feasible level (see Table 4). Intermediate scores were assigned relative to these endpoints. Table 6 shows the scores. Below, the general rationales for the scores are described on a criterion-by-criterion basis.

Criterion 1.1.1 Driving Skills - This criterion refers to the construct validity of a test. High construct validity depends upon how direct (as opposed to inferential) a measure is, this primarily being a function of the salience of the behavior or skill to the driving task. A test scored high if it directly assessed some behavior explicitly involved in driving (such as vehicle control or some other fundamental element of the driving task), such that failure to perform on the test would comprise direct evidence of inability to perform the driving task. Test tended to receive

TABLE 6

COMPARATIVE ASSESSMENTS OF BEHAVIOR TESTS ON EVALUATION CRITERIA

Criteria	Auditory Frequency Differential Threshold	Auditory Intensity Differential Threshold	Auditory Threshold	Autokinesis	Brightness Sensitivity	Color Discrimination	Critical Flicker Fusion	Delayed Auditory Feedback	Duration of Archimedes Spiral Afterimage	Dynamic Visual Acuity	Glare Recovery
1.0											
1.1											
1.1.1	0	0	20	0	20	5	10	10	5	70	35
1.1.2	2	2	5	2	5	20	5	25	2	30	15
1.1.3	70	50	70	70	70	50	60	60	70	80	70
1.1.4	5	10	10	10	10	10	10	15	10	15	10
1.1.5	15	10	10	5	25	45	85	75	10	90	30
1.2											
1.2.1	70	65	70	5	70	70	60	35	35	85	50
1.2.2	90	80	80	50	40	75	90	80	90	90	35
1.2.3	60	60	60	40	80	60	90	35	80	90	90
1.2.4	75	75	75	80	75	95	75	65	80	75	80
1.2.5	80	80	80	90	80	80	90	10	90	75	80
2.0											
2.1											
2.1.1	10	30	60	80	50	60	50	50	70	40	80
2.1.2	15	15	15	15	15	60	30	15	15	25	30
2.1.3	100	100	100	70	65	80	65	40	80	100	100
2.1.4	80	80	80	40	80	90	60	60	80	45	70
2.1.5	60	60	60	0	60	75	20	25	30	60	15
2.2											
2.2.1	95	95	95	40	95	100	100	80	100	100	60
2.2.2	90	90	90	40	90	65	90	90	95	80	90
2.3	60	5	30	5	30	10	30	30	5	30	30

TABLE 6 (Continued)

COMPARATIVE ASSESSMENTS OF BEHAVIOR TESTS ON EVALUATION CRITERIA

Criteria	Nystagmus	Peripheral Vision	Static Visual Acuity	Two-Point Tactile Discrimination	Visual Field	Color Naming	Distance Estimation	Time Estimation	Arithmetic	Code Substitution	Complex Counting	Counting Backwards
1.0												
1.1												
1.1.1	0	70	60	10	45	15	60	65	15	15	15	15
1.1.2	45	40	30	2	25	10	25	30	2	2	5	5
1.1.3	70	60	70	60	70	70	25	40	65	60	10	10
1.1.4	75	20	20	10	15	20	15	10	20	15	50	50
1.1.5	45	20	70	5	20	15	25	20	30	65	45	45
1.2												
1.2.1	70	80	90	25	80	45	40	30	40	60	35	30
1.2.2	90	90	90	70	90	80	40	65	60	80	90	90
1.2.3	100	90	90	80	90	80	70	30	80	80	80	80
1.2.4	100	90	80	100	90	95	95	90	100	95	100	100
1.2.5	100	90	80	80	80	30	50	50	10	10	15	20
2.0												
2.1												
2.1.1	20	85	50	40	85	50	80	50	80	30	70	80
2.1.2	90	40	30	90	40	90	95	80	100	75	100	100
2.1.3	5	70	100	50	70	70	40	80	30	100	40	40
2.1.4	20	80	50	80	80	90	70	90	75	90	75	75
2.1.5	30	30	60	10	30	55	70	70	50	40	75	75
2.2												
2.2.1	70	80	100	60	80	90	100	100	100	100	100	100
2.2.2	10	25	85	20	25	80	95	90	90	50	90	90
2.3	60	20	10	60	10	20	40	40	10	30	35	35

TABLE 6 (Continued)

COMPARATIVE ASSESSMENTS OF BEHAVIOR TESTS ON EVALUATION CRITERIA

Criteria	Digit Memory	Serial Performance	Sorting	Stroup Test	Tongue Twisters	Choice Reaction Time	Compensatory Tracking	Critical Tracking	Pursuit Tracking	Simple Reaction Time	Tracing	Bender Drawing Test
1.0												
1.1												
1.1.1	15	5	10	2	5	55	65	60	60	50	30	5
1.1.2	2	5	10	5	5	25	35	40	30	20	10	0
1.1.3	60	60	60	5	80	10	35	35	35	40	60	0
1.1.4	15	10	40	35	5	15	30	30	30	20	50	55
1.1.5	40	30	40	20	5	30	70	65	75	35	30	5
1.2												
1.2.1	20	15	20	40	60	30	55	80	65	40	25	60
1.2.2	50	60	60	80	40	50	65	65	50	50	80	85
1.2.3	70	50	50	80	95	35	40	40	30	35	70	90
1.2.4	100	80	95	95	0	80	50	45	65	85	95	95
1.2.5	10	40	60	20	20	90	80	80	80	90	50	65
2.0												
2.1												
2.1.1	75	40	30	40	80	50	30	20	40	55	30	10
2.1.2	100	25	50	70	90	25	10	10	10	25	57	70
2.1.3	100	100	90	90	0	100	100	100	100	100	70	0
2.1.4	85	90	70	90	70	80	60	60	65	80	90	10
2.1.5	40	50	70	55	40	55	50	50	50	55	20	10
2.2												
2.2.1	100	95	95	100	100	95	95	95	95	95	100	95
2.2.2	90	80	60	75	90	95	95	95	95	95	50	50
2.3	20	20	20	39	20	55	65	65	65	55	10	5

TABLE 6 (Continued)

COMPARATIVE ASSESSMENTS OF BEHAVIOR TESTS ON EVALUATION CRITERIA

Criteria	Digit-Symbol Substitution Task	Letter Cancellation (Bourdon)	Minnesota Spatial Relations Test	Trail Making Test	Finger Count	Finger-to Finger	Finger-to Nose	Grip Strength	Modified Romberg Test	One-Hole Test	One-Leg Stand	Picking up Objects
1.0												
1.1												
1.1.1	25	15	20	15	15	15	15	5	20	25	20	25
1.1.2	5	5	10	10	15	35	25	2	30	10	30	30
1.1.3	40	75	60	60	80	80	60	70	80	80	80	80
1.1.4	25	30	40	25	60	60	65	5	65	15	60	70
1.1.5	65	50	50	40	10	20	20	40	30	20	30	10
1.2												
1.2.1	35	40	60	20	40	50	50	60	50	50	50	15
1.2.2	75	80	60	80	90	90	70	90	60	80	50	50
1.2.3	50	70	70	70	80	90	90	40	90	70	90	80
1.2.4	80	95	95	95	100	100	100	95	100	90	100	95
1.2.5	30	40	60	25	70	80	80	80	80	70	80	80
2.0												
2.1												
2.1.1	50	70	30	0	90	90	90	90	80	80	85	80
2.1.2	25	75	50	75	100	100	100	80	100	80	100	95
2.1.3	100	70	80	70	40	30	30	100	30	90	30	10
2.1.4	90	90	70	80	80	75	75	90	75	90	75	75
2.1.5	40	20	65	20	75	75	75	40	75	70	65	60
2.2												
2.2.1	100	100	95	100	100	75	60	90	45	95	40	50
2.2.2	50	50	80	50	90	95	95	85	95	95	95	95
2.3	30	30	40	30	40	40	10	20	40	20	40	30

TABLE 6 (Continued)

COMPARATIVE ASSESSMENTS OF BEHAVIOR TESTS ON EVALUATION CRITERIA

Criteria	Romberg Test	Tapping	Walk and Turn	Wobble Board	Divided Attention	Field Dependence	Vigilance
1.0							
1.1							
1.1.1	20	10	40	20	80	15	75
1.1.2	30	2	30	35	45	40	50
1.1.3	80	80	80	80	40	80	60
1.1.4	70	15	40	70	15	40	20
1.1.5	30	50	25	30	80	50	80
1.2							
1.2.1	50	75	50	60	65	65	40
1.2.2	60	90	70	60	50	70	40
1.2.3	90	40	90	90	40	60	35
1.2.4	100	80	100	50	40	75	80
1.2.5	80	70	80	80	80	20	80
2.0							
2.1							
2.1.1	85	80	75	70	50	80	0
2.1.2	100	60	100	20	10	25	25
2.1.3	30	100	25	100	100	100	100
2.1.4	75	85	75	70	55	80	80
2.1.5	75	40	60	75	40	60	50
2.2							
2.2.1	50	100	55	35	90	100	90
2.2.2	95	95	95	90	95	90	60
2.3	20	20	40	40	75	40	60

intermediate scores if they measure a skill or behavior indicative of the capacity or capability to perform certain elements of the driving task, even if that capacity/capability cannot be directly assessed in actual driving. Tests that cannot be related directly to driving performance or to driver capability/capacity, or that relate only to isolated skills tangentially related to the driving task received low scores.

Criterion 1.1.2 Predicting Accidents - Although no behavioral (or chemical) test has high absolute predictability with regard to accidents, tests may differ in implied relative probabilities of accidents. A test may have high predictive validity because it directly measures a critical skill or capacity/capability in regard to the driving tasks (i.e., because it has high construct validity as defined by Criterion 1.1.1). However, a test need not have high construct validity to have predictive validity. A test not obviously related to the driving task may have high predictive validity because it detects gross, exceedingly high levels of impairment such that the human is unlikely to be able to perform any task (driving or other) competently. A test receives a low score if it does not provide a direct measure of a driving skill or capability/capacity crucial to accident avoidance, or if it is not designed to detect grossly high levels of impairment.

Criterion 1.1.3 Identifying conditions not measured - Two general considerations give rise to scores on this criterion. First, tests scored highly if explicit evidence exists that they do not measure certain impairing agents. Second, tests scored poorly if explicit evidence exists that shows ambiguous measurement of impairing agents. Across studies, tests that sometimes discriminate an agent and other times fail to discriminate received relatively lower scores. Based on these considerations, tests of visual function generally scored high, since (a) amphetamines do not improve performance on them, and (b) they are insensitive to marijuana. Psychomotor tasks such as one-leg-stand, wobble board, and walk and turn are not impaired by amphetamines. Tracking and reaction time tasks have shown mixed results from several drugs.

Criterion 1.1.4 Single-test interpretation - A behavioral test will typically be administered to a driver suspected of being impaired, and the officer will not have baseline data. A test scored high on this criterion if it validly discriminates among an impaired individual and "his/her non-impaired self" and the normal population of drivers. A test scored low on this criterion if the performance measures for "impaired" individuals tend to fall within the performance range of the normal population.

Criterion 1.1.5 Multiple causes - The more impairing agents a test has been shown to detect, the higher the score it received on this criterion. On the psychologically complex tasks, the ability to link particular error patterns with particular sources of impairment yielded high scores. Some tasks fail to distriminate any cause of impairment. Other tasks are so simple that results fail to indicate impairment from multiple causes. Some sensory-perceptual tasks are susceptible only to impairment from a single cause, e.g., alcohol. Certain tasks have shown the ability to discriminate between amphetamines and barbituates (better-than-normal versus worse-than-normal performance), whereas other tasks are complex enough to have several error patterns which provide evidence as to the specific cause of impairment.

Criterion 1.2.1 Test-retest reliability - Many factors can contribute to poor test-retest reliability, but the criterion refers only to aspects of the test itself. Some tests (e.g., arithmetic) are inherently very reliable. Tests scored lower on this criterion for any of the following reasons: (1) the test has a time restriction; (2) the test has either a floor or ceiling with regard to score; (3) the scoring of the test is predicated on variability rather than location; or (4) the determination of test score involves indirect procedures (e.g., psychophysical staircase methods).

Criterion 1.2.2 Ambient conditions - A perfect score of 100 on this criterion reflects total independence from surrounding environmental/physical conditions. Tests were scored lower if they are affected by external light sources, rain or wind, or slope of the area in which the test is administered. Tests that require a high degree of concentration were scored lower, since they are affected by noise as well as by other interfering conditions. Sensory tests were rated high because they require no conscious attention. Cognitive tasks generally were given high ratings, with some downgrading for those requiring high levels of concentration. Motor tests are generally quite sensitive to ambient environmental conditions. Tests that are performed in the open also received low ratings. Certain tests with unusual requirements were also rated low: vigilance and divided attention, because of the amount of attention required, and distance estimation, because of the probable need for light to see distant targets. Some tests scored relatively high because it was assumed that headphones or other devices would provide sources of ambient condition control.

Criterion 1.2.3 Non-alterable performance - Tasks on which learning can significantly improve performance were rated low. Generally, such tasks included motor tasks, particularly those requiring fine motor performance. Tests on which poor performance can be faked with little chance of detection were rated relatively low, as were tests where good performance could be faked. Sensory tests generally received high scores because of their insensitivity to conscious control by the subject.

Criterion 1.2.4 Reliability of equipment - A test received a 100 on this criterion if absolutely no equipment is required for administration, or if the required equipment is independent of operator error and cannot be miscalibrated. Tests requiring progressively more complex equipment, or tests requiring multiple pieces of equipment, were scored progressively lower.

Criterion 1.2.5 Demographic populations - It is assumed that norms for tests will be established for (1) different age groups and (2) males versus females. Since it is impractical to establish norms for each separate language, ethnic, and educational-level group, reliability across these groups is desirable. Tests tended to score high if they are free of cultural biases of all forms; tests tended to score low if they involve verbal, intelligence/achievement, or other skills that may be culturally related.

Criterion 2.1.1 Duration of test - A test was scored 100 if it required no time to administer (i.e., the officer need merely observe the behavior in the course of his/her normal duties). A test was scored 0 if its administration (including instructions) required five minutes or more. On tests for which multiple measures are required to ensure reliability (e.g., reaction time), duration was assumed to be sufficient to produce reliability. Scores also take into account any time required to set up equipment or score tests.

Criterion 2.1.2 Equipment - A test received a score of 100 if no equipment is required for administration, and 0 if the required equipment is approximately the size of a medium piece of luggage or larger. In all cases, equipment was assumed to shield external conditions to the extent possible and provide automated testing and scoring, if feasible.

Criterion 2.1.3 Ease of scoring - Tests which have automatic scoring or can be easily modified to provide automatic scoring received ratings of 100. Low ratings were given to those tests that required a high degree of subjective interpretation on the part of the test administrator.

Criterion 2.1.4 Training - A score of 100 denoted a test that is entirely self-explanatory. A score of 0 was given to tests requiring extensive training, testing, and operator certification. Tests were scored lower that required subjective judgment (administrator) of performance since they would require the officer to have previous relevant experience. A test scored lower if training was required for use of equipment. Tests were not penalized where it could be assumed that much of the administration and scoring could be automated.

Criterion 2.1.5 Cooperation of subject - Tests were rated low on this criterion if they required a subject to endure uncomfortable circumstances, either because of the test itself (e.g., glare recovery) or because of the equipment that would be required to successfully perform the test (e.g., a hood for some tests). In addition, low ratings were given to complex cognitive tasks whose instructions might be difficult to follow. High ratings were given to the simple cognitive tests requiring few and easy instructions, and to the simple motor tests involving gross motor skills.

Criterion 2.2.1 Safety to suspect - Scores of 100 reflected tests with no risk to the suspect, while scores of 0 indicated fairly significant physical or mental (e.g., stress, anxiety) risk to the subject. Tests were scored lower if any of the sensory nodalities were inhibited during administration. This consideration includes tests requiring headphones, eye covers, and the closure of eyes. Tests scored lower that required intense concentration, or that could easily provoke mental stress in other ways (e.g., frustration, confusion). Susceptibility to physical injury also caused tests to be scored low (e.g., wobble board).

Criterion 2.2.2 Safety to officer - Perfect scores of 100 indicated no threat was present in either administration or scoring of test. Tests were scored lower that required intense concentration by the administering officer; such tests reduce awareness of what the suspect is doing. In addition, tests were scored lower if they either limited officer access to weapons or provided such access to the testee. Tests scored highly if they failed to require the administrator to be in close physical proximity with the testee. The effect of these factors on scoring was mediated by test duration; the longer the test, the less safe it generally is.

Criterion 2.3 Diagnostic value - Since some tests fail to discriminate any sources of impairment, they fail to have any diagnostic value whatsoever. Moderate scores reflected the ability to identify one or two particular sources of impairment. Even tests that appear to differentiate among sources of impairment are variable in their ability to identify certain classes of impairment and were, thus, downgraded somewhat.

3.5 Quantitative Evaluation Results

Overall evaluation results were obtained by weighting the score on each criterion by the normalized weight assigned to that criterion and summing across criteria. Thus, because two different sets of weights were used to represent the perspectives of users and researchers, there are two somewhat different sets of results. In addition to the overall results, results on each of the major criteria (e.g., safety) could also be calculated. These quantitative results are presented in Table 7.

TABLE 7

QUANTITATIVE EVALUATION RESULTS OF BEHAVIORAL TESTS

	Auditory Frequency Differential Threshold	Auditory Intensity Differential Threshold	Auditory Threshold	Autokinesis	Brightness Sensitivity	Color Discrimination	Critical Flicker Fusion	Delayed Auditory Feedback	Duration of Archimedes Spiral Afterimage	Dynamic Visual Acuity	Glare Recovery
USERS											
1.1 Validity	11	10	14	11	18	22	29	32	12	44	22
1.2 Reliability	72	70	70	52	68	73	85	52	81	86	71
2.1 Use	39	49	64	53	54	68	44	41	59	52	63
2.2 Safety	91	91	91	40	91	71	92	88	96	83	85
2.3 Diagnosis	60	5	30	5	30	10	30	30	5	30	30
OVERALL	55	56	64	42	60	61	61	54	64	65	63
RESEARCHERS											
1.1 Validity	8	7	17	8	19	16	23	25	10	54	27
1.2 Reliability	72	71	72	63	78	72	87	25	82	82	80
2.1 Use	43	51	63	42	56	70	39	38	53	54	53
2.2 Safety	92	92	92	40	92	77	93	87	97	87	80
2.3 Diagnosis	60	5	30	5	30	10	30	30	5	30	30
OVERALL	43	43	50	35	51	50	52	33	48	64	53

TABLE 7 (Continued)

QUANTITATIVE EVALUATION RESULTS OF BEHAVIORAL TESTS

	Nystagmus	Peripheral Vision	Static Visual Acuity	Two-Point Tactile Discrimination	Visual Field	Color Naming	Distance Estimation	Time Estimation	Arithmetic	Code Substitution	Complex Counting	Counting Backwards
USERS												
1.1 Validity	55	32	40	11	25	20	24	24	22	27	34	34
1.2 Reliability	96	90	87	79	89	80	65	53	74	79	82	82
2.1 Use	24	67	58	39	67	58	73	63	67	47	69	74
2.2 Safety	20	34	87	27	34	82	96	92	92	58	92	92
2.3 Diagnosis	60	20	10	60	10	20	40	40	10	30	35	35
OVERALL	42	56	67	40	55	61	69	62	66	53	71	73
RESEARCHERS												
1.1 Validity	36	45	47	11	32	19	37	38	20	23	27	27
1.2 Reliability	97	89	85	76	84	53	57	43	42	44	45	48
2.1 Use	25	59	58	33	59	58	72	64	63	46	70	74
2.2 Safety	30	43	90	33	43	83	97	93	93	67	93	93
2.3 Diagnosis	60	20	10	60	10	20	40	40	10	30	35	35
OVERALL	52	61	63	39	54	43	56	50	42	38	48	49

TABLE 7 (Continued)

QUANTITATIVE EVALUATION RESULTS OF BEHAVIORAL TESTS

	Digit Memory	Serial Performance	Sorting	Stroup Test	Tongue Twisters	Choice Reaction Time	Compensatory Tracking	Critical Tracking	Pursuit Tracking	Simple Reaction Time	Tracing	Bender Drawing Test
USERS												
1.1 Validity	21	16	33	21	10	24	44	43	43	28	37	25
1.2 Reliability	67	57	62	79	54	51	51	51	47	52	75	87
2.1 Use	71	52	51	55	60	58	45	40	50	61	38	12
2.2 Safety	92	82	66	79	92	95	95	95	95	95	58	57
2.3 Diagnosis	20	20	20	25	20	55	65	65	65	55	10	5
OVERALL	67	55	53	60	58	61	59	57	60	63	50	40
RESEARCHERS												
1.1 Validity	20	13	25	14	9	35	51	49	49	36	35	18
1.2 Reliability	37	44	54	48	20	63	62	64	59	64	58	75
2.1 Use	65	52	55	55	56	58	46	42	50	60	34	11
2.2 Safety	93	85	72	83	93	95	95	95	95	95	67	65
2.3 Diagnosis	20	20	20	25	20	55	65	65	65	55	10	5
OVERALL	41	37	44	39	30	54	57	56	56	55	43	37

TABLE 7 (Continued)

QUANTITATIVE EVALUATION RESULTS OF BEHAVIORAL TESTS

	Digit-Symbol Substitution Task	Letter Cancellation (Bourdon)	Minnesota Spatial Relations Test	Trail Making Test	Finger Count	Finger-to-Finger	Finger-to-Nose	Grip Strength	Modified Romberg Test	One-Hole Test	One-Leg Stand	Picking up Objects
USERS												
1.1 Validity	31	31	36	27	38	42	42	17	47	21	45	46
1.2 Reliability	62	75	72	74	85	90	85	68	82	76	80	73
2.1 Use	55	59	49	23	80	78	78	78	76	79	73	66
2.2 Safety	58	58	82	58	92	92	89	86	87	95	86	87
2.3 Diagnosis	30	30	40	30	40	40	10	20	40	20	40	30
OVERALL	53	57	60	42	76	77	75	67	74	72	72	69
RESEARCHERS												
1.1 Validity	29	26	31	23	30	34	33	13	38	23	37	39
1.2 Reliability	41	54	65	45	73	82	82	64	81	69	81	74
2.1 Use	52	51	52	22	79	78	78	70	73	77	71	65
2.2 Safety	67	67	85	67	93	88	83	87	78	95	77	80
2.3 Diagnosis	30	30	40	30	40	40	10	20	40	20	40	30
OVERALL	41	43	50	33	59	63	61	47	62	55	61	59

TABLE 7 (Continued)

QUANTITATIVE EVALUATION RESULTS OF BEHAVIORAL TESTS

	Romberg Test	Tapping	Walk and Turn	Wobble Board	Divided Attention	Field Dependence	Vigilance
USERS							
1.1 Validity	50	24	38	51	43	42	47
1.2 Reliability	82	65	85	72	45	64	48
2.1 Use	76	72	66	73	52	75	32
2.2 Safety	87	96	88	81	94	92	65
2.3 Diagnosis	20	20	40	40	75	40	60
OVERALL	75	68	70	70	61	71	46
RESEARCHERS							
1.1 Validity	40	20	39	40	56	33	57
1.2 Reliability	81	60	82	81	62	41	59
2.1 Use	76	65	65	73	50	72	35
2.2 Safety	80	97	82	72	93	93	70
2.3 Diagnosis	20	20	40	40	75	40	60
OVERALL	63	48	61	62	60	49	54

These scores should not be interpreted too precisely: differences of a few points may not be very meaningful. Rather, they provide a rough indication of the overall potential of the tests. Thus, for example, although the finger-to-finger test received the highest overall score for law enforcement officers, it should not necessarily be considered better than other tests with scores that are nearly as high including complex counting, counting backward, field dependence, finger count, walk and turn, wobble board, romberg, modified romberg, one leg stand, one hole test, and finger-to-nose; all of which received scores of 70 or higher. Tests scoring relatively low (below 50) for law enforcement officers include the Bender drawing test, two point tactile discrimination, vigilance, autokinesis, nystagmus, and trail making. The reasons for high and low scores can be found by examining Tables 5, 6, and 7. Vigilance, for example, receives a low score overall because it does poorly with respect to reliability and ease of use, two criteria which receive relatively high weight. Tests that scored well overall generally received high scores on all criteria, or at least the relatively important ones.

The results are, of course, somewhat different from the researchers' perspective. For researchers, tests generally received somewhat lower scores, because of the stronger emphasis on validity which has been demonstrated to be high for very few tests. Note, however, that this is not surprising in light of the conditions described in Section 2.0. Recall also that only moderate levels of validity do not preclude the valid use of behavioral tests for impairment.

For researchers the top tests included dynamic visual acuity, static visual acuity, walk and turn, wobble board, finger-to-finger, finger-to-nose, romberg, modified romberg, one leg stand, simple peripheral vision, and divided attention. Although there is considerable overlap with the most promising tests from the law enforcement officers' perspective, particularly for the tests of gross motor performance, some tests are different. This occurred primarily because the tests in the top group for researchers scored relatively well on criteria that the researchers considered most important (i.e., validity and reliability) and less well on ease of use and safety which were relatively more important from the officers' perspective. The opposite is true, of course, for tests in the top group from the officers' perspective that are not in this group for the researchers. Table 8 provides a summary of the most promising tests from both perspectives.

Similarly, the researchers' least promising tests overlapped some, but not completely with the officers' low group. The overlap includes the Bender drawing test, two point tactile discrimination, autokinesis, and trail making. Additional tests with very little potential for the researchers are the Stroup test, tongue twisters, code substitution, serial performance, and delayed auditory feedback.

In considering the development of a battery of tests, a concept discussed more fully in the following section, some attention might be given to including tests requiring different types of behavior. The only major type of behavior identified in Table 1 that is not represented in the list of most promising tests in Table 8 is perceptual-motor behavior. The two reaction time tests (simple and choice) and the three tracking tests (pursuit, compensatory, and critical) are the most promising tests from both perspectives. For the law enforcement officers, the reaction time tests are slightly preferred with simple reaction time, choice reaction time, pursuit tracking, compensatory tracking, and critical tracking being ranked 23.5, 28, 32.5, 35, and 37.5 respectively. The researchers prefer the tracking tasks slightly with overall rankings for the same five tests of 18.5, 21, 16, 14, and 16 respectively.

TABLE 8

MOST PROMISING TESTS FROM TWO PERSPECTIVES

Test	Rank Order	
	Law Enforcement Officers	Researchers
Dynamic Visual Acuity	20	1
Peripheral Vision	39.5	8.5
Static Visual Acuity	17	3
Complex Counting	9.5	34
Counting Backwards	6	31.5
Finger Count	2	12.5
Finger-to-Finger	1	3
Finger-to-Nose	3.5	8.5
Modified Romberg	5	5.5
One-Hole Test	7.5	18.5
One-Leg Stand	7.5	8.5
Romberg	3.5	3
Walk and Turn	11.5	8.5
Wobble Board	11.5	5.5
Divided Attention	28	11
Field Dependence	9.5	31.5

4.0 DISCUSSION AND RECOMMENDATIONS

This assessment and evaluation of behavioral tests of driving impairment indicates that they should be pursued as a means of identifying impaired drivers in order to remove such drivers from the road. Several of these tests do relatively well with respect to the three primary areas of concern: technical performance (primarily validity), practical implementation, and legal issues and constraints.

The tests are not differentiated by the legal analysis (see Appendix). Rather, it suggests that any reasonable behavioral test provides useful evidence, primarily in support of chemical evidence. Where chemical tests are not possible, evidence from behavioral tests becomes even more important. This, of course, is within the context of current relevant statutes. As we note below, some change may be possible.

Practical criteria (e.g., ease of use and safety) serve primarily to distinguish among various tests with respect to their potential rather than exclude any particular behavioral tests. These criteria are, however, important in determining which tests to pursue, particularly because they are relatively important to the law enforcement officers who would administer the tests.

Perhaps the most questionable aspect of behavioral tests is their technical performance, specifically their validity. Validity can take many forms. Some argue for prediction of accidents, others for relation to driving performance. In any case, it is very difficult to determine. Most likely, for reasons discussed in Section 2.0, it is also moderate, at best, for any particular test. But even low validity does not preclude the use of behavioral tests. In particular, a battery of behavioral tests, each of which have only low validity, can be used to construct an impairment index that can have rather high validity. One of the purposes of this evaluation is to suggest tests for inclusion in such a battery.

Although the evaluation considered only individual tests, it does suggest some considerations in constructing a battery of tests, and also indicates some additional research that could support the use of such a battery. Among the considerations in constructing the battery are:

- use tests that were evaluated highly,
- use tests tapping different driving behaviors or behaviors in general, and
- use tests that detect impairment produced by as many causes as possible.

Thus, from among the top-rated tests, dynamic visual acuity, counting backwards, finger-to-finger, and romberg might be selected for a test battery. These tests cover a range of behaviors including sensory-perceptual, cognitive, and two types of motor performance. Only the dynamic visual acuity test requires any equipment. Consideration should also be given to including a divided attention task that was rated highly with respect to validity, but was downgraded because it is relatively more difficult to use and has lower reliability. It could be composed of a tracking and a choice reaction time task, thus covering perceptual-motor performance which is not measured by any of the other four tests.

Research concerning whether or not performance on these tests is changed as the result of impairment produced by a wide range of agents is lacking. Few agents, other than alcohol and to a certain extent marijuana have been extensively tested for their effects on behavior.

A research program is needed to test which impairing agents can be detected by these behavioral tests. In addition, epidemiological studies are needed so that if certain agents are not detected by the battery of tests, it will be known whether the risk produced by these agents is sufficient to justify additional research into behavioral tests that can detect impairment caused by these agents.

This suggests a long-term research program that is needed. In the near-term, however, one question needs to be addressed before much further attention is given to behavioral tests for detecting impairment. We need to have a somewhat better understanding of how judges and juries view behavioral tests and their results. For example, would they view the behavioral tests suggested for the battery as requiring skills associated with driving so that poor performance on the test would be taken as strong evidence that the person was impaired with respect to driving ability. That is, the behavioral tests should provide strong evidence without the need for an expert witness. The legal analysis conducted as part of this study (see Appendix) indicates that judges and juries are generally willing to accept behavioral test results as evidence, but we need to know whether or not there are differences among behavioral tests in this regard. Behavioral tests whose results are less acceptable in court, although they can be used to identify drivers who should be removed from the road, lose much of their usefulness.

This discussion of behavioral test results as evidence presumes no change in relevant state statutes. If behavioral tests, and particularly a specific battery of tests, are further demonstrated to be a valid and practical means for identifying impaired drivers, model legislation could be developed that would guide states toward changing statutes to make better use as evidence of the results of behavioral tests. There is no philosophical reason, for example, why results from behavioral tests cannot have the same evidentiary status as results from chemical tests. Practical considerations such as the reliability of behavioral measures would, of course, also be a factor. We are not, however, arguing here for behavioral tests to have the same legal status as chemical tests given the current state-of-the-art. Rather, we are suggesting that a very broad range of legal possibilities exist, and extreme possibilities should not be rejected out of hand without some careful consideration.

A final aspect of the use of behavioral tests to detect impaired drivers that deserves further attention is the implementation of such tests. Sophisticated implementation strategies can enhance the use of such tests and reduce the time and effort required to conduct them. For example, the tests in a battery could be administered at roadside, with additional tests at the station if the roadside tests provided sufficient indication of impairment. This approach would be particularly useful if the roadside tests involved little or no equipment, and the tests administered at the station were those requiring special equipment. This sequential approach would also be useful if videotaping of performance on behavioral tests proved to be useful evidence.

Finally, it is worth discussing here how our evaluations differ somewhat from the results of related previous studies. Burns and Moskowitz (1977) in their study of several behavioral tests concluded that nystagmus, the one-leg stand, and the walk and turn tests were to be preferred for detecting alcohol-induced impairment. Our analysis found the one-leg stand and the walk and turn to be evaluated rather highly. Nystagmus, on the other hand, was not a highly-rated test. There are several reasons for this difference. First, Burns and Moskowitz evaluated tests with respect to the relationship between performance on the test and blood-alcohol concentration (BAC). A close relationship between these two variables does not necessarily imply a close relationship between performance on the nystagmus test and driving performance, or between test performance and accidents. Specifically, it is not apparent that performance on the nystagmus test reflects any skills related to driving. In addition, examining a driver for nystagmus may be difficult operationally and somewhat unsafe. Scoring is quite subjective and would require careful training for the test administrator. It also requires the administrator to be in very close proximity to the subject and to concentrate closely on the subject's eyes which makes it difficult to observe other motions by the subject, e.g., with the hands.

Other studies have produced different results for various reasons. For example, Edwards et al. (1973), studying the effects of alcohol on performance of behavioral tests and driving, concluded that two tracking tasks (compensatory and critical) and a choice reaction time task were the most promising for detecting impairment. They did not investigate, however, tests such as the romberg, finger-to-finger, walk and turn, etc. that do not require any equipment, since they were looking for a test to be used as an interlock device in automobiles.

These two studies, like most extensive investigations of behavioral tests, looked only at the effects of alcohol. Although alcohol-produced impairment obviously played a major role in our assessment of behavioral tests, we expanded the focus to include other impairing agents, thereby possibly changing the evaluations of behavioral tests from what they might have been considering only alcohol.

REFERENCES

1. Adams, A. J., Brown, B., Flom, M. C., Jones, R. T., and Jampolsky, A. Alcohol and marihauna effects on static visual acuity. American Journal of Optometry and Physiological Optics, 1975, 52, 729-735.

2. Adams, A. J., Brown, B., Haegerstrom-Portnoy, G., Flom, M. C., and Jones, R. T. Evidence for acute effects of alcohol and marijuana on color discrimination. Perception and Psychophysics, 1976, 20, 119-124.

3. Adams, A. J., Brown, B., Haegerstrom-Portnoy, G., Flom, M. C., and Jones, R. T. Marijuana, alcohol, and combined drug effects on the time course of glare recovery. Psychopharmacology, 1978, 56, 81-86.

3a. American Institutes for Research. Development of a standardized battery of performance tests for the assessment of noise stress effects. Silver Spring, MD: NASA CR-2149, January 1973.

4. Aschan, G., Bergstedt, M., Goldberg, Z., and Laurell, L. Positional nystagmus in man during and after alcohol intoxications. Quarterly Journal of Studies on Alcohol, 1956, 17, 381-408.

5. Asmussen, E., and Boje, O. The effect of alcohol and some drugs on the capacity for work. Acta Physiologica Scandinavia, 1948, 15, 109-113.

6. Austin, G. A., Sterling-Smith, R. S., Macari, M. A., and Lettieri, D. J. Drug users and driving behaviors. National Institute on Drug Abuse, June 1977.

7. Ayers, J. L., Templer, D. I., Ruff, C. F., and Barthlow, V. L. Trail making test improvement in abstinent alcoholics. Journal of Studies on Alcohol, 1978, 39, 1627-1629.

8. Barrett, G. V., and Thornton, D. L. Relationship between perceptual style and driver reaction to an emergency situation. Journal of Applied Psychology, 1968, 52, 169-176.

9. Barrett, G. V., Thornton, C. L., and Cabe, P. A. Relation between embedded figures test performance and simulator behavior. Journal of Applied Psychology, 1969, 53, 253-254.

10. Baugh, R., and Calvert, R. T. The effect of dipherhydramine alone and in combination with ethanol on histamine skin response and mental performance. European Journal of Clinical Pharmacology, 1977, 12, 201-204.

11. Bayliss, P. F. C., and Duncan, S. M. The clinical pharmacology of viloxazine hydrochloride. British Journal of Clinical Pharmacology, 1974, 1, 431-437.

12. Begbie, G. H. The effects of alcohol and of varying amounts of visual information on a balancing test. Ergonomics, 1966, 9, 325-333.

13. Beinstein, M. E., Richards, A. B., Hughes, F. W., and Forney, R. B. Optokinetic nystagmus under the influence of d-amphetamine and alcohol. In R. N. Harger (Ed.) Alcohol and traffic safety. Bloomington, IN: Indiana University Press, 1966.

14. Bell, R. W., Dickie, D. S., Stewart-Jones, J., and Turner, P. Lorazepan on visuo-motor coordination and visual function in man. Journal of Pharmacy and Pharmacology, 1973, 25, 87-88.

15. Bernhard, C. G., and Skoglund, C. R. Selective suppression with ethyl alcohol of inhibition in the optic nerve and of the negative component P III of the electroretinogram. Acta Physiologica Scandinavia, 1941, 2, 10-21.

16. Betts, T. A., and Blake, A. The psychotropic effects of atenolol in normal subjects: Preliminary findings. Postgraduate Medical Journal, 1977, 53, 151-156.

17. Biehl, B. The effects of two tranquillizers on driving performance as measured in the normal driving task. Paper presented at 18th Congress of International Association of Applied Psychology, Montreal, 1974.

18. Blum, B., Stern, M. H., and Melville, K. I. A comparative evaluation of the action of depressant and stimulant drugs on human performance. Psychopharmacologia, 1964, 6, 173-177.

19. Boersma, F. J., Muir, W., Wilson, K., and Barnham, R. Eye movements during embedded figure tasks. Perceptual and Motor Skills, 1969, 28, 271-274.

20. Booher, H. R. Effects of visual and auditory impairment in driving performance. Human Factors, 1978, 20, 307-320.

21. Borkenstein, R. F., Crowther, R. F., Shumate, R. P., Ziel, W. B., and Zylman, R. The role of the drinking driver in traffic accidents. Bloomington, Indiana: Indiana University, Department of Police Administration, 1964.

22. Borland, R. G., and Nicholson, A. N. Comparison of the effects of two benzodiazepines and pertobarbitone sodium or human performance. British Journal of Clinical Pharmacology, 1975, 2, 9-17. (a)

23. Borland, R. G., and Nicholson, A. N. Immediate effects on human performance of a 1,5 benzodiazepine (clobazam) compared with the 1,4 benzodiazepines chlordiazepoxide hydrochloride and diazepam. British Journal of Clinical Pharmacology, 1975, 2, 215-222. (b)

24. Brown, B., Adams, A. J., Haegerstrom-Portnoy, G., Jones, R. T., and Flom, M. C. Effects of alcohol and marijuana on dynamic visual acuity: I. Threshold measurements. Perception and Psychophysics, 1975, 18, 441-446.

25. Burford, B., French, I., and LeBlanc, A. E. The combined effects of alcohol and common psychoactive drugs: I. Studies on human pursuit tracking capability. In S. Israelstam and S. Lambert (Eds.) Alcohol, drugs, and traffic safety proceedings. Proceedings of the Sixth International Conference on Alcohol, Drugs, and Traffic Safety, Toronto, September 8-13, 1974. Toronto, Canada: Addiction Research Foundation of Ontario, 1975.

26. Burns, M., and Moskowitz, H. Psychophysical tests for DWI arrest: Training manual. Southern California Research Institute, Contract No. DOT-HS-8-01970.

27. Burns, M., and Moskowitz, H. Psychophysical tests for DWI arrest. DOT HS-802 424, Southern California Research Institute, June 1977.

28. Burns, M., and Sharma, S. Marihuana effects on driving: Performance and personality. In D. F. Huieke (Ed.) Proceedings of the 19th Conference of the American Association for Automotive Medicine, San Diego, California, November 20-22, 1975. Lake Bluff, IL: The Association, 1975.

29. Bye, C., Clubley, A. W., and Peck, A. W. Drowsiness, impaired performance and tricyclic antidepressant drugs. British Journal of Clinical Pharmacology, 1978, 6, 155-161.

30. Bye, C., Munro-Faure, A. D., Peck, A. W., and Young, P. A. A comparison of the effects of 1-benzylpiperazine and dexamphetamine on human performance tests. European Journal of Clinical Pharmacology, 1973, 6, 163-169.

31. Caldwell, D. F., Myers, S. A., Domino, E. F., and Merriam, P. E. Auditory and visual threshold effects of marihuana in man. Perceptual and Motor Skills, 1969, 29, 755-759.

32. Callaway, E. The influence of amobarbital (amylobarbitone) and methamphetamine on the focus of attention. Journal of Mental Science, 1959, 105, 382-392.

33. Casswell, S., and Marks, D. Cannabis induced impairment of performance of a divided attention task. Nature, 1973, 241, 60-61.

34. Cheser, G. B., Franks, H. M., Jackson, D. M., Starmer, G. A., and Teo, R. K. C. Ethanol and delta-a-tetrahydrocannabinol: Interactive effects on human perceptual, cognitive and motor functions: II. Medical Journal of Australia, 1977, 1, 478-481.

35. Clark, L. D., and Nakashima, E. N. Experimental studies of marihuana. American Journal of Psychiatry, 1968, 125, 379-384.

36. Clarke, C. H., and Nicholson, A. N. Immediate and residual effects in man of the metabolites of diazepam. British Journal of Clinical Pharmacology, 1978, 6, 325-331.

37. Clayton, A. B., Betts, T. A., and Mackay, G. M. A study of the effects of certain tranquillizers and small amounts of alcohol on driving performance. European Journal of Toxicology, 1972, 5, 254-256.

38. Clayton, A. B., Mackay, G. M., and Betts, T. A. The effects of certain tranquillizers and alcohol upon kinetic visual acuity. Proceedings of 16th Conference American Association for Automotive Medicine, 1972, 199-215.

39. Colson, Z. W. The effect of alcohol on vision. Journal of American Medical Association, 1940, 115, 1525-1527.

40. Crancer, A., Dille, J. M., Delay, J. C., Wallace, J. E., and Haykin, M. D. Comparison of the effects of marihuana and alcohol on simulated driving performance. Science, 1969, 164, 851-854.

41. Crow, J. W., Lain, P., Bochner, F., Shoeman, D. W., and Azarnoff, D. L. Glutethimide and 4-OH glutethimide: Pharmacokinetics and effect on performance in man. Clinical Pharmacology and Therapeutics, 1977, 22, 458-464.

42. Currie, L. The perception of danger in a simulated driving task. Ergonomics, 1969, 12, 841-849.

43. Dalton, W. S., Martz, R., Lemberger, L., Rodda, B. E., and Forney, R. B. Effects of marihuana combined with secobarbitol. Clinical Pharmacology and Therapeutics, 1975, 18, 298-304.

44. Day, E. S., Jones, S., Stewart-Jones, J., and Turner, P. Clemastine on hand-eye coordination and visual function. Journal of Clinical Pharmacology (and New Drugs), 1972, 12, 240.

45. Dee, H. L., and Van Allen, M. W. Psychomotor testing as an aid in the recognition of cerebral lesions. Neurology, 1972, 22, 845-848.

46. Demers, R. G., and Heninger, G. R. Visual-motor performance during lithium treatment. A preliminary report. Journal of Clinical Pharmacology (and New Drugs), 1971, 11, 274-279.

47. Dickens, D. W., Lader, M. H., and Steinberg, H. Differential effects of two amphetamine-barbituate mixtures in man. British Journal of Pharmacology, 1965, 24, 14-23.

48. Donelson, A. C., Marks, M. E., Jones, R. K., and Joscelyn, K. B. Drug research methodology, Volume One: The alcohol-highway safety experience and its applicability to other drugs. University of Michigan, Highway Safety Research Institute, DOT-HS-805-374, 1980.

48a. Duncan, J. Divided attention: The whole is more than the sum of its parts. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 216-228.

49. Edwards, W., Goodman, B., and Snapper, K. J. Breath, blood, and behavioral indices of impairment (Technical Report to General Motors). Ann Arbor: University of Michigan, Institute of Science and Technology, Engineering Psychology Laboratory, July 1971.

50. Edwards, W. How to use multiattribute utility measurement for social decision making. IEEE Transactions on Systems, Man, and Cybernetics, 1977, 7, 326-340.

51. Edwards, W., Lathrop, J., Seaver, D., Seghers, R., and Lehman, R. Development and evaluation of behavioral tests of driver impairment. Highway Safety Research Institute, University of Michigan, August 1973.

52. Enzer, N., Simonson, E., and Ballard, G. The effect of small doses of alcohol on the central nervous system. American Journal of Clinical Pathology, 1944, 14, 333-341.

53. Evans, B. S., Martz, R. M., Brown, D. J., Rodda, B. E., Kiplinger, B. F., Lemberger, L., and Forney, R. B. Impairment of performance with low doses of marihuana. Clinical Pharmacology and Therapeutics, 1970, 14, 936-940.

54. Farris, R., Malone, T. B., and Lilliefors, H. A comparison of alcohol involvement in exposed and injured drivers, Phases I and II. National Highway Traffic Safety Administration, Technical Report DOT-HS-801-826, 1976.

55. Forney, R. B., and Hughes, F. W. Delayed auditory feedback and ethanol: Effect on verbal and arithmetical performance. Journal of Psychology, 1961, 52, 185-192.

56. Forney, R. B., and Hughes, F. W. Meprobamate, ethanol or meprobamate-ethanol combinations on performance of human subjects under delayed audio feedback (DAF). Journal of Psychology, 1964, 57, 431-436.

57. Frankenhaueser, M. Drug effects on emotions: Relevance to driving accidents. In S. Israelstam and S. Lambert (Eds.) Alcohol, drugs, and traffic safety proceedings. Proceedings of the Sixth International Conference on Alcohol, Drugs, and Traffic Safety, Toronto, September 8-13, 1974. Toronto, Canada: Addiction Research Foundation of Ontario, 1975.

58. Franks, H. M., Hensley, V. R., Starmer, G. A., and Teo, R. K. C. The interaction between ethanol and antihistamines, I: Dexchlorpheniramine. Medical Journal of Australia, 1978, 1, 449-452.

59. Fregly, A. R., Bergstedt, M., and Graybiel, A. Relationships between blood alcohol, positional alcohol nystagmus and postural equilibrium. Quarterly Journal of Studies on Alcohol, 1967, 28, 11-21.

60. Gamberale, F., Ornwall, G., and Hultengren, M. Exposure to xylene and ethylbenzene: III. Effects on central nervous functions. Scandinavian Journal of Work Environment and Health, 1978, 4, 204-211.

61. Garfield, J. M., Garfield, F. B., and Sampson, J. Effects of nitrous oxide on decision-strategy and sustained attention. Psychopharmacologia, 1975, 42, 5-10.

62. Gershon, S., Holmberg, G., Matteson, E., Matteson, N., and Marshall, A. Imipramine hydrochloride: Its effects upon clinical, autonomic, and psychological functions. Archives of General Psychiatry, 1962, 6, 96-101.

63. Goldberg, L. Behavioral and physiological effects of alcohol on man. Psychosomatic Medicine, 1966, 28, 570-595.

64. Goldstein, L. G. Research on human variables in safe motor vehicle operation: A correlational summary of predictor variables and criterion measures. Washington, D.C. George Washington University, Driver Behavior Research Project, June 1961.

65. Goodenough, D. R. A review of individual differences in field dependence as a factor in auto safety. Human Factors, 1976, 18, 53-62.

66. Goodnow, R. E., Beecher, H. K., Brazier, M. A. B., Mosteller, F., and Tagiuri, R. Physiological performance following a hypnotic dose of a barbituate. Journal of Pharmacology and Experimental Therapeutics, 1951, 102, 55-61.

67. Gordon, N. B. Reaction-times of methadone treated ex-heroin addicts. Psychopharmacologia, 1970, 16, 337-344.

68. Griffis, L. C., Bright, T. P., Cerimele, B. J., and Forney, R. B. Combined effects of fermetozole and ethanol. Clinical Pharmacology and Therapeutics, 1978, 24, 350-353.

69. Haffner, J. F. W. Mental and psychomotor effects of diazepam and ethanol. Acta Pharmacologia et Toxicologia, 1973, 32, 161-178.

70. Harano, R. M. Relationship of field dependence and motor vehicle accident involvement. Perceptual and Motor Skills, 1970, 31, 272-274.

71. Hart, J., Hill, H. M., Bye, C. E., Wilkinson, R. T., and Peck, A. W. The effects of low doses of amylobarbitone sodium and diazepam on human performance. British Journal of Clinical Pharmacology, 1976, 3, 289-298.

72. Hartocollis, P. Drunkeness and suggestion: An experiment with intravenous alcohol. Quarterly Journal of Studies on Alcohol, 1962, 23, 376-389.

73. Harvey, P. G., Clayton, A. B., and Betts, T. A. The effects of four antihypertensive agents on the Stroop Colour-Word Test in normal male volunteer subjects. Psychopharmacology, 1977, 54, 133-138.

74. Harvey, P. G., Clayton, A. B., and Betts, T. A. The effects of two antidepressant agents on performance on the Stroop Colour-Word Test. British Journal of Clinical Pharmacology, 1978, 5, 305-312.

75. Hebbard, F. W., and Fischer, R. Effects of psilocybin, LSD, and mescaline on small, involuntary eye movements. Psychopharmacologia, 1966, 9, 146-156.

76. Hebbelinck, M. The effects of a small dose of ethyl alcohol on certain basic components of human physical performance: II. The effect on neuromuscular performances. Archives of International Pharmacodynamics, 1963, 143, 247-257.

77. Hecht, A. Drugs and driving. In Food and Drug Administration Consumer, September 1978, p. 17-19.

78. Hindmarch, I. Some aspects of the effects of clobozam on human psychomotor performance. British Journal of Clinical Pharmacology, 1979, 7, 77-82.

79. Holmberg, G., and William-Olsson, V. The effect of benzquinamide in comparison with chlordiazepoxide and placebo, on performance in some psychological tests. Psychopharmacologia, 1963, 4, 402-417.

80. Holzman, P., Levy, D. L., Uhlenhuth, E. H., Proctor, L. R., and Freedman, D. X. Smooth-pursuit eye movements and diazepam, CPZ, and secobarbital. Psychopharmacologia, 1975, 44, 111-115.

81. Howells, D. E. Nystagmus as a physical sign in alcohol intoxication. British Medical Journal, 1956, 1, 1405-1406.

82. Hughes, F. W., and Forney, R. B. Comparative effect of three antihistaminics and ethanol on mental and motor performance. Clinical Pharmacology and Therapeutics, 1964, 5, 414-421.

83. Hughes, F. W., and Forney, R. B. Delayed audiofeedback for induction of anxiety: Effect of nortriptyline, ethanol or nortriptyline-ethanol combinations on performance with DAF. Journal of the American Medical Association, 1963, 185, 556-558.

84. Hughes, F. W., Forney, R. B., and Gates, P. W. Performance in human subjects under delayed auditory feedback after alcohol, a tranquilizer (benzquinamide) or benzquinamide-alcohol combinations. Journal of Psychology, 1963, 55, 25-32.

85. Hughes, F. W., Forney, R. B., and Richards, A. B. Comparative effect in human subjects of chlordiazepoxide, diazepam, and placebo on mental and physical performance. Clinical Pharmacology and Therapeutics, 1965, 6, 139-145.

86. Hurst, P. M., and Bagley, S. K. Acute adaptation to the effects of alcohol.. Quarterly Journal on Studies of Alcohol, 1972, 33, 358-378.

87. Hutchinson, H. C., Tuchtie, M., Gray, K. G., and Steinberg, D. A study of the effects of alcohol on mental functions. Canadian Psychiatric Association Journal, 1964, 9, 33-42.

88. Huffman, W. J., Florio, A. E., Payne, J. L., and Boys, F. E. The influence of two selected tranquillizers on driving skills. American Journal of Psychiatry, 1963, 119, 885-886.

89. Idestrom, C., and Cadenius, B. Chlordiazepoxide, dipiperon, and amobarbitol dose effect studies on human beings. Psychopharmacologia, 1963, 4, 235-246.

90. Idestrom, C. M., and Cadenius, B. Time relations of the effects of alcohol compared to placebo. Psychopharmacologia, 1968, 13, 189-200.

91. Ikeda, H., and Granger, G. W. Action of alcohol on visual and retinal responses to intermittent illumination. In Alcohol and road traffic - proceedings 3rd int. conference. London: British Medical Association, 1963.

92. Jaattela, A., Mannisto, P., Paatero, H., and Tuomisto, J. The effects of diazepam or diphenhydramine on healthy human subjects. Psychopharmacologia, 1971, 21, 202-211.

93. Jellinek, E. M., and McFarland, R. A. Analysis of psychological experiments on the effects of alcohol. Quarterly Journal of Studies on Alcohol, 1940, 1, 272-371.

94. Joscelyn, K. B., and Donelson, A. C. Drugs research methodology. Volume III. The detection and quantification of drugs of interest in body fluids from drivers. University of Michigan, Highway Safety Research Institute. Technical Report DOT-HS-805-375, 1980. (a)

95. Joscelyn, K. B., and Donelson, A. C. Drug research methodology. Volume V. The study of drug effects on skills related to driving. University of Michigan, Highway Safety Research Institute. Technical Report DOT-HS-7-01530, 1980. (b)

96. Joscelyn, K. B., and Donelson, A. C. Drugs and driving: A selected bibliography. Supplement One. National Highway Traffic Safety Administration Technical Report DOT-HS-803-879, 1979.

97. Joscelyn, K. B., Jones, R. K., Maickel, R. P., and Donelson, A. C. Drugs and driving: Information needs and research requirements. National Highway Traffic Safety Administration Technical Report Number DOT-HS-804-774, 1979.

98. Joscelyn, K. B., and Maickel, R. P. Report on an international symposium on drugs and driving (DOT HS-802-187). Washington, D.C.: U.S. Department of Transportation, National Highway Traffic Safety Administration, January 1977.

99. Kaplan, H. L., Forney, R. B., Hughes, F. W., and Richards, R. B. Comparative effects in human subjects of three hypnotics and placebo on mental and motor performance. Archives Internationales de Pharmacodynamie et de Therapie, 1968, 174, 181-191.

100. Kaplan, R. J., Lathrop, J., and Edwards, W. Measurement of risk-related and other dimensions of alcohol-induced impairment. Ann Arbor, MI: University of Michigan, Highway Safety Research Institute, December 1976.

101. Kataja, M., Penttila, A, and Tenhu, M. Combining the blood alcohol and clinical examinations for estimating the influence of alcohol. Blutalkohol, 1975, 12, 108-115.

102. Keeney, R. L., and Raiffa, H. Decisions with multiple objectives: Preferences and value tradeoffs. New York: Wiley, 1976.

103. Kelley, E. L., Miller, T. G., Marquis, D. G., Gerard, R. W., and Uhr, L. Personality differences and continued meprobamate and proclorperazine administration. Archives of Neurology and Psychiatry, 1958, 80, 241-246.

104. Kielholz, P., Hobi, V., Ladewig, D., Miest, P., and Richter, R. An experimental investigation about the effect of cannabis on car driving behaviour. Pharmakopsychiatrie Neuro-Psychopharmakologie, 1973, 6, 91-103.

105. King, A. R. Tunnel vision. Quarterly Journal of Studies on Alcohol, 1943, 4, 362-367.

106. Kiplinger, G. F., Manno, J. E., Rodda, B. E., and Forney, R. B. Dose-response analysis of the effects of tetrahydrocannabinol in man. Clinical Pharmacology and Therapeutics, 1971, 12, 650-657.

107. Klein, A. W., Davis, J. H., and Blackbourne, B. D. Marihuana and automobile crashes. Journal of Drug Issues, 1971, 1, 18-26.

108. Kleinknecht, R. A., and Donaldson, D. A review of the effects of diazepam on cognitive and psychomotor performance. Journal of Nervous and Mental Disease, 1975, 161, 399-414.

109. Kornetsky, C., and Orzack, M. H. Research note on some of the critical factors on the dissimilar effects of chlorpromazine and secobarbital on the digit symbol substitution and continuous performance tests. Psychopharmacologia, 1974, 6, 79-86.

110. Kulshreshtha, V. K., Gupta, P. P., Turner, P., and Wadsworth, J. Some clinical pharmacological studies with terfenadine, a new antihistamine drug. British Journal of Clinical Pharmacology, 1978, 6, 25-29.

111. Landaver, A. A., Laurie, W., and Milner, G. The effect of benzoctamine and alcohol on motor-skills used in car driving. Forensic Science, 1973, 2, 275-283.

112. Landaver, A. A., and Milner, G. Antihistamines, alone and together with alcohol, in relation to driving safety. Journal of Forensic Medicine, 1971, 18, 127-139.

113. Landaver, A. A., and Milner, G. Desipramine and imipramine, alone and together with alcohol in relation to driving safety. Pharmakopsychiatrie Neuro-Psychopharmakologie, 1971, 4, 265-275.

114. Landaver, A. A., Milner, G., and Patman, J. Alcohol and amitriptyline effects on skills related to driving behavior. Science, 1969, 163, 1467-1468.

115. Landaver, A. A., Pocock, D. A., and Prott, F. W. Effects of atenolol and propramolol on human performance and subjective feelings. Psychopharmacology, 1979, 60, 211-215.

116. Landaver, A. A., Pocock, D. A., and Prott, F. W. The effects of medazepam and alcohol on cognitive and motor skills used in driving. Psychopharmocologia, 1974, 37, 158-168.

117. Langolf, G. D., Chaffin, D. B., Henderson, R., and Whittle, H. P. Evaluation of workers exposed to elemental mercury using quantitative tests of tremor and neuromuscular functions. American Industrial Hygiene Association Journal, 1978, 39, 976-984.

118. Large, A. T. W., Wayte, G., and Turner, P. Promethazine on hand-eye co-ordination and visual function. Journal of Pharmaceutical Pharmacology, 1971, 23, 134-135.

119. Laties,.V. G., and Weiss, B. Effects of alcohol on timing behavior. Journal of Comparative and Physiological Psychology, 1962, 55, 85-91.

120. Lawton, M. P., and Cahn, B. The effects of diazepam (valium) and alcohol on psychomotor performance. Journal of Nervous and Mental Diseases, 1963, 136, 550-554.

121. Legg, J. F., and Stiff, M. P. Drug-related test patterns of depressed patients. Psychopharmacology, 1976, 50, 205-210.

122. Lewis, E. G., Dustman, R. E., and Beck, E. C. The effect of alcohol on sensory phenomena and cognitive and motor tasks. Quarterly Journal of Studies on Alcohol, 1969, 30, 618-633.

123. Liljequist, R., Linnoila, M., and Mattila, M. J. Effect of diazepam and chlorpromazine on memory functions in man. European Journal of Clinical Pharmacology, 1978, 13, 339-343.

124. Linnoila, M. Drug effects on psychomotor skills related to driving: Interaction of atropine, glycopyrrhonium and alcohol. European Journal of Clinical Pharmacology, 1973, 6, 107-112.

125. Linnoila, M. Interaction of isoniazid and alcohol on psychomotor skills related to driving. In 7th Annual Meeting of the Finnish Medical Societies, Helsinki, January 8-9, 1973 (abstract).

126. Linnoila, M., Erwin, C. W., Cleveland, W. P., Logue, P. E., and Gentry, W. D. Effects of alcohol on psychomotor performance of men and women. Journal of Studies on Alcohol, 1978, 39, 745-758.

127. Linnoila, M., and Mattila, M. J. Drug interaction on psychomotor skills related to driving: Diazepam and alcohol. European Journal of Clinical Pharmacology, 1973, 5, 186-194.

128. Linnoila, M., and Mattila, M. J. Combined effects of diazepam and alcohol on skills related to driving. In 6th Annual Meeting of the Finnish Medical Society, Helsinki, March 23-25, 1972 (abstract).

129. Linnoila, M., Saario, F., and Maki, M. Effect of treatment with diazepam or lithium and alcohol on psychomotor skills related to driving. European Journal of Clinical Pharmacology, 1974, 7, 337-342.

130. Lubin, R. A. Influences of alcohol, interpersonal feedback, and drinking experience upon performance and judgment. Perceptual and Motor Skills, 1979, 48, 95-104.

131. Macavoy, M. G., and Marks, D. F. Divided attention performance of cannabis users and non-users following cannabis and alcohol. Psychopharmacology, 1975, 44, 147-152.

132. Manno, J. E., Kiplinger, G. F., Haine, S. E., Bennett, I. F., and Forney, R. B. Comparative effects of smoking marihuana or placebo on human motor and mental performance. Clinical Pharmacology and Therapeutics, 1970, 11, 808-815.

133. Manno, J. E., Kiplinger, G. F., Scholz, N., and Forney, R. B. The influence of alcohol and marihuana on motor and mental performance. Clinical Pharmacology and Therapeutics, 1971, 12, 202-211.

134. Mattila, M. J., Palva, E., Seppala, T., and Ostrovskaya, R. V. Actions and interactions with alcohol of drugs on psychomotor skills: Comparison of diazepam and gammahydroxybutyric acid. Archives Internationales de Pharmacodynamie et de Therapie, 1978, 234, 236-246.

135. Mattila, M. J., Palva, E., Seppala, T., and Saario, I. Effects of trithiozine on psychomotor skills related to driving: A comparison with diazepam and interactions with alcohol. Current Therapeutic Research, 1977, 22, 875-884.

136. McCarroll, J. R., and Haddon, W., Jr. A controlled study of fatal automobile accidents in New York City. Journal of Chronic Diseases, 1962, 15, 811-26.

137. McKnight, A. A., and Adams, B. B. Driver education task analysis Volume II: Task analysis methods (DOT HS 800 368). Alexandria, VA: Human Resources Research Organization, November 1970.

138. Melikian, L. The effect of meprobamate on the performance of normal subjects on selected psychological tests. Journal of General Psychology, 1961, 65, 33-38.

139. Meyer, R. E., Pillard, R. C., Shapiro, L. M., and Mirin, S. M. Administration of marijuana to heavy and casual marijuana users. American Journal of Psychiatry, 1971, 128, 90-96.

140. Mihal, W. L., and Barrett, G. V. Individual differences in perceptual informational processing and their relation to automobile accident involvement. Journal of Applied Psychology, 1976, 61, 229-233.

141. Miller, J. G. Objective measurements of the effects of drugs on driver behavior. Journal of American Medical Association, 1962, 179, 940-943.

142. Milner, G., and Landaver, A. A. Alcohol, thioridazine, and chlorpromazine effects on skills related to driving behavior. British Journal of Psychiatry, 1971, 118, 351-352 (abstract).

143. Milner, G., and Landaver, A. A. Haloperidol and diazepam, alone and together with alcohol, in relation to driving safety. Blutalkohol, 1973, 10, 247-254.

144. Milstein, S. L., MacCannel, K., Karr, G., and Clark, S. Marihuana-produced impairments in coordination. Journal of Nervous and Mental Disease, 1975, 161, 26-31.

145. Mirsky, A. F., and Kornetsky, C. On the dissimilar effects of drugs on the digit symbol substitution and continous performance tests: A review and preliminary integration of behavioral and physiological evidence. Psychopharmacologia, 1964, 5, 161-177.

146. Monton, J. G. The effect of sedative drugs on human performance. In I. R. Johnston (Ed.) Proceedings of the Seventh International Conference on Alcohol, Drugs, and Traffic Safety, Canberra, Australia, 1979.

147. Morland, J. Combined effects of diazepam and ethanol on mental and psychomotor functions. Acta Pharmacologia et Toxicologia, 1974, 34, 5-15.

148. Moskowitz, H. Validity of driving simulator studies for predicting drug effects in real driving situations. In: Israelstam, S., and Lambert, S. (eds.) Alcohol, Drugs, and Traffic Safety. Proceedings of the Sixth International Conference on Alcohol, Drugs, and Traffic Safety. Toronto, September 8-13, 1974. Toronto, Canada: Addiction Research Foundation of Ontario, 1975, pp. 295-304.

149. Moskowitz, H., and DePry, D. Differential effect of alcohol on auditory vigilance and divided-attention tasks. Quarterly Journal of Studies on Alcohol, 1968, 29, 54-63.

150. Moskowitz, H., Hulbert, S., and McGlothlin, W. H. Marihuana: Effects on simulated driving performance. Accident Analysis and Prevention, 1976, 8, 45-50.

151. Moskowitz, H., and McGlothlin, W. H. Effects of marihuana on auditory signal detection. Psychopharmacologia, 1974, 40, 137-145.

152. Moskowitz, H., and Sharma, S. Effects of alcohol on peripheral vision as a function of attention. Human Factors, 1974, 16, 174.

153. Moskowitz, H., Sharma, S., and Shapero, M. A comparison of the effects of marijuana and alcohol on visual functions. In Current research in marijuana. New York: Academic Press, 1972.

154. Moskowitz, H., Sharma, S., and Ziedman, K. The effect of a single acute dose of diazepam on driving-related skills performance. In Proceedings of the 23rd Conference of the American Association for Automotive Medicine. Morton Grove, IL: AAAM, 1979.

155. Mould, G. P., Curry, S. H., and Binns, T. B. Interaction of glutethimide and phenobarbitone with ethanol in man. Journal of Pharmaceutical Pharmacology, 1972, 24, 894-899.

156. Muller-Oerlinghausen, B., Bauer, H., Girke, W., Kanowski, S., and Goncalves, N. Impairment of vigilance and performance under lithium treatment. Pharmakopsychiatrie Neuro-Psychopharmakologie, 1977, 10, 67-78.

157. Murphree, H. B., Price, L. M., and Greenberg, L. A. Effect of congeners in alcoholic beverages on the incidence of nystagmus. Quarterly Journal of Studies on Alcohol, 1966, 27, 201-213.

158. Nash, H. Alcohol and caffeine: A study of their psychological effects. Springfield, IL: Charles C. Thomas, 1962.

158a. National Highway Traffic Safety Administration. Marijuana, other drugs, and their relation to highway safety. Report to Congress, 1979.

159. Newman, H., and Fletcher, E. The effect of alcohol on vision. American Journal of Medical Science, 1941, 202, 723-731.

160. Newman, H. W., and Newman, E. J. Failure of dexedrine and caffeine as practical antagonists of the depressant effect of ethyl alcohol in man. Quarterly Journal of Studies on Alcohol, 1956, 17, 406-410.

161. Olson, P. L. Aspects of driving performance as a function of field dependence. Journal of Applied Psychology, 1974, 59, 192-196.

162. Patman, J., Landaver, A. A., and Milner, G. The combined effects of alcohol and amitriptyline on skills similar to motorcar driving. Medical Journal of Australia, 1969, 2, 946-949.

163. Palva, E. S., and Linnoila, M. Effect of active metabolites of chlordiazepoxide and diazepam, alone or in combination with alcohol, on psychomotor skills related to driving. European Journal of Clinical Pharmacology, 1978, 13, 345-350.

164. Palva, E. S., Linnoila, M., and Mattila, M. J. Effect of active metabolites of chlordiazepoxide and diazepam, alone or in combination with alochol, on psychomotor skills related to driving. In J. J. Mattila (Ed.) Alochol, drugs, and driving. (Vol. II) Basel: S. Krager, 1976.

165. Pearson, R. G., and Neal, G. L. Operator performance as a function of drug, hypoxia, individual and task factors. Aerospace Medicine, 1970, 41, 154-158.

166. Peek, A. W., Adams, R., Bye, C., and Wilkinson, R. T. Residual effects of hypnotic drugs: Evidence for individual differences on vigilance. Psychopharmacology, 1976, 47, 213-216.

167. Penttila, A., Lenti, H., and Lonnquist, J. Nystagmus and disturbances in psychomotor functions induced by psychotropic drug therapy. Psychiatria Fennica, 1974, 315-326.

168. Perrine, M. W., (ed). Alcohol, drugs, and driving. National Highway Traffic Safety Administration Technical Report DOT-HS-801-096, 1974.

169. Perrine, M. W., Waller, J. A., and Harris, L. S. Alcohol and highway Safety: Behavioral and medical aspects. National Highway Traffic Safety Administration Technical Report DOT-HS-800-599, 1971.

170. Peters, H. B. Changes in color fields occasioned by experimentally induced intoxication. Journal of Applied Psychology, 1942, 26, 692-701.

171. Pihkanen, T. A. On static atactic functional disorders caused by alcohol. Quarterly Journal of Studies on Alcohol, 1957, 18, 183-189.

172. Pihkanen, T., and Kauko, O. The effects of alcohol on the perception of musical stimuli. Annals Medicinae et Biologiae Fennice, 1962, 40, 275-285.

173. Prag, J. J. The chemical and clinical diagnosis of "driving under the influence of alcohol" and the use of the chemical tests in traffic law enforcement. South African Journal of Clinical Science, 1953, 4, 289-325.

174. Putz, V. R. The effects of carbon monoxide on dual-task performance. Human Factors, 1979, 21, 13-24.

175. Quarton, G. C., and Talland, G. A. The effects of methamphetamine and pentobarbital on two measures of attention. Psychopharmacologia, 1962, 3, 66-71.

176. Radford, L. M., Chaney, E. F., O'Leary, M. R., and O'Leary, D. E. Screening for cognitive impairment among inpatients. Journal of Clinical Psychiatry, 1978, 39, 712-715.

177. Rafaelson, O. J., Bech, P., and Rafaelson, L. Simulated car driving influenced by cannabis and alcohol. Pharmakopsychiatrie Neuro- Psychopharmakologie, 1973, 6, 71-83.

178. Roback, G. S., Krasno, L. R., and Ivy, A. C. Effect of analeptic drugs on the somnifacient effect of second and antihistaminics as measured by the flicker fusion threshold. Journal of Applied Physiology, 1952, 4, 566-574.

179. Roth, W. T., Tinklenberg, J. R., Whitaker, C. A., Darley, C. F., Kopell, B. S., and Hollister, L. E. The effect of marihuana on tracking task performance. Psychopharmacologia, 1973, 33, 259-265.

180. Rundell, O. H., and Williams, H. L. Alcohol and speed-accuracy tradeoff. Human Factors, 1979, 21, 433-443.

181. Rutschmann, J., and Rubenstein, L. Time estimation, knowledge of results and drug effects. Journal of Psychiatric Research, 1966, 4, 107-114.

182. Salvendy, G., and McCabe, G. P., Jr. Marijuana and human performance. Human Factors, 1975, 17, 229-235.

182a. Salvendy, G., and Seymour, W. D. Prediction and development of industrial work performance. New York: Wiley, 1973.

182b. Salvendy, G., and Seymour, W. D. The manual of the one-hole test. Lafayette, Indiana: Lafayette Instrument Company, 1972.

183. Schwartzberg, A. Z., and Van de Castle, R. W. A controlled clinical study of chlordiazepoxide. American Journal of Psychiatry, 1961, 117, 922-923.

184. Schwinn, R., Hill, S. Y., Goodwin, D. W., and Powell, B. Marihuana and critical flicker fusion. Journal of Nervous and Mental Disease, 1974, 158, 142-144.

185. Seppala, T., Linnoila, M., Elonen, E., Mattila, M. J., and Maki, M. Effect of tricyclic antidepressants and alcohol on psychomotor skills related to driving. Clinical Pharmacology and Therapeutics, 1975, 17, 515-522.

186. Seppala, T., Saario, I., and Mattila, M. S. Two weeks' treatment with chlorpromazine, thioridazine, sulpiride, or bromazepam: Actions and interactions with alcohol on psychomotor skills related to driving. In J. J. Mattila (Ed.) Alcohol, drugs, and driving (Vol. II). Basel: S. Krager, 1976.

187. Sharma, S. Marihuana effects on a critical tracking task. In D. F. Hueike (Ed.) Proceedings of the 19th Conference of the American Association for Automotive Medicine, San Diego, Calif., Nov. 20-22, 1975. Lake Bluff, IL: The Association, 1975.

188. Sharma, S., and Moskowitz, H. Effect of marihuana on the visual autokinetic phenomenon. Perceptual and Motor Skills, 1972, 35, 891-894.

189. Sharma, S., and Moskowitz, H. Effect of two levels of attention demand on vigilance performance under marihuana. Perceptual and Motor Skills, 1974, 38, 967-970.

190. Sharma, S. Sedatives. In Drugs and Driving, ed. R. E. Willete, pp. 61-72. NIDA Research Monograph II. Washington, D.C.: U.S. Government Printing Office, 1976.

191. Shinar, D., McDowell, E. D., Rackoff, N. J., and Rockwell, T. H. Field dependence and driver visual search behavior. Human Factors, 1978, 20, 553-559.

192. Smith, H. W. Pharmacology of alcohol and alcohol-drug combinations. In: Proceedings of the 4th International Conference on Alcohol and Road Traffic. Bloomington, Indiana: Indiana University, 1966, pp. 26-34.

193. Snapper, K. J. Effects of alcohol on psychomotor skill and decision-making in a driving task (Technical Report). Ann Arbor: University of Michigan, Institute of Science and Technology, Engineering Psychology Laboratory, 1973.

194. Snapper, K. J., O'Connor, M. F., and Einhorn, H. J. Social indications: A new method for indexing quality (Technical Report 74-7). Washington, D.C.: George Washington University, Social Research Group, October 1974.

195. Sommerbeck, K. W., Theilgaard, A., Rasmussen, K. E., Lohren, V., Gram, L., and Wolff, K. Valproate sodium: Evaluation of so-called psychotropic effect. A controlled study. Epilepsia, 1977, 18, 159-167.

196. Takala, M., Siro, E., and Toivainen, Y. Intellectual functions and dexterity during hangover. Quarterly Journal of Studies on Alcohol, 1958, 19, 1-29.

197. Thorndike, R. Personnel selection: Test and measurement techniques. New York: Wiley, 1949.

198. Turner, P. Critical flicker frequency and centrally-acting drugs. British Journal of Ophthalmology, 1968, 52, 245-249.

199. Tyrer, P. J., and Lader, M. H. Physiological and psychological effects of propranolol and diazepam in anxiety. British Journal of Clinical Pharmacology, 1974, 1, 379-385.

200. Uhr, L., Pollard, J. C., and Miller, J. G. Behavioral effect of chronic administration of psychoactive drugs to anxious subjects. Psychopharmacologia, 1959, 1, 150-160.

201. Veldkamp, M. B., Donelson, A. C., and Joscelyn, K. B. Drugs and Driving: A Selected Bibliography. Supplement Two. Ann Arbor: University of Michigan, Highway Safety Research Institute, 1980.

201a. Veldkamp, M. B., Donelson, A. C., and Joscelyn, K. B. Drugs and Driving: A Selected Bibliography. Supplement Three. Ann Arbor: University of Michigan, Highway Safety Research Institute, June 1980.

202. Vernon, R. J., and Ferguson, R. K. Effects of trichloroethylene on visual-motor performance. Archives of Environmental Health, 1969, 18, 894-900.

203. Verriest, G., and LaPlasse, D. New data concerning the influence of ethyl alcohol on human visual thresholds. Experimental Eye Research, 1965, 4, 95-101.

204. Vivkari, M., Linnoila, M., and Aalto, V. Efficacy and side effects of flurazepam, fosazepam, and nitrazepam as sleeping aids in psychogeriatric patients. Acta Psychiatricia Scandinavice, 1978, 57, 27-35.

205. Vuchinich, R. E., and Sobell, M. B. Empirical separation of physiologic and expected effects of alcohol on complex perceptual motor performance. Psychopharmacology, 1978, 60, 81-85.

206. Waller, D. F., Li, L. K., Hall, R. G., and Stutts, J. C. Driver performance tests: Their role and potential. Contract DOT-HS-7-01698, March, 1978.

207. Wallgren, H., and Barry, H., III. Actions of alcohol. Amsterdam: Elsevier, 1970.

208. Wambsganss, V. E., and Bredenkamp, J. An experimental psychological study of the effects of halo peridol. Arzneimittel-Forschung, 1968, 18, 238-241.

209. Webb, E. W., Campbell, D. T., Schwartz, R. D., and Sechrest, L. Unobtrusive measures: Nonreactive research in the social sciences. Chicago, Illinois: Rand McNally, 1966.

210. Weil, A. T., Zinberg, N. E., and Nelson, J. M. Clinical and psychological effects of marihuana in man. Science, 1968, 162, 1234-1242.

211. Wetherell, A. Individual and group effects of 10 mg diazepam on drivers' ability, confidence and willingness to act in a gap-judging task. Psychopharmacology, 1979, 63, 259-267.

212. Wilkinson, R. T., and Colquhoun, W. P. Interaction of alcohol with incentive and with sleep deprivation. Journal of Experimental Psychology, 1968, 76, 623-629.

213. Willette, R. E., ed. Drugs and driving. National Institute of Drug Abuse Research, Monograph II. U.S. Department of Health, Education, and Welfare, Publication No. (ADM) 77-432, 1977.

214. Wilson, L., Taylor, J. D., Nash, C. W., and Cameron, D. F. The combined effects of ethanol and amphetamine sulfate on performance of human subjects. Canadian Medical Association Journal, 1966, 94, 478-484.

215. Wittenborn, J. R., Flaherty, C. F., McGough, W. E., Bossange, K. A., and Nash, R. J. A comparison of the effect of imipramine, nomifensine, and placebo on the psychomotor performance of normal males. Psychopharmacology, 1976, 51, 85-90.

216. Wittenborn, J. R., Flaherty, C. F., McGough, W. E., and Nash, R. J. Psychomotor changes during initial day of benzodiazepine medication. British Journal of Clinical Pharmacology, 1979, 7, 69-76.

217. Wright, G., Randell, P., and Shephard, R. J. Carbon monoxide and driving skills. Archives of Environmental Health, 1973, 27, 349-354.

218. Zabik, J. E. Measurement of drug effects on behavior. Chapter VI in Report on an international symposium on drugs and driving (DOT HS-802-187). Washington, D.C.: U.S. Department of Transportation, National Highway Traffic Safety Administration, January 1977.

219. Zimmerman-Tansella, C., Tansella, M., and Lader, M. The effects of chlordesmethyldiazepam on behavioral performance and subjective judgment in normal subjects. Journal of Clinical Pharmacology, 1976, 16, 481-488.

220. Zirkle, G. A., King, P. D., McAtee, O. B., and Van Dyke. Effects of chlorpromazine and alcohol on coordination and judgment. Journal of American Medical Association, 1959, 171, 1496-1499.

221. Zirkle, G. A., McAtee, O. B., King, P. D., and Van Dyke. Meprobamate and small amounts of alcohol. Journal of American Medical Association, 1960, 173, 1823-1825.

222. Zylman, R. Age is more important than alcohol in the collision-involvement of young and old drivers. Journal of Traffic Safety Education, 1972, 20(1), 7-9.

223. Zylman, R. A critical evaluation of the literature on "alcohol involvement" in highway deaths. Accident Analysis and Prevention, 1974, 6, 163-204.

APPENDIX

ANALYSIS OF LEGAL ISSUES

Elliot R. Levine, Esq.

A.1 Summary

A.1.1 Legal relevance of behavioral tests. Behavioral tests do have legal relevance; results of such tests may be admitted into evidence to show that the defendant was under the influence of drugs. Courts find such tests useful because they provide a police officer with a multitude of behavioral signs which help support the charge that the defendant was under the influence of drugs.

A.1.2 Evidential role of behavioral tests. Behavioral tests are "circumstantial" rather than "direct" evidence of the presence of drugs in the system. "Circumstantial" evidence is valuable evidence for the fact finder (whether judge or jury) and may under some circumstances suffice standing alone to obtain a conviction. However, where under the facts and circumstances of a case the trier (evaluator) of fact may make a reasonable inference of innocence, circumstantial evidence may not by itself be enough to obtain a conviction. Thus, behavioral test results, while helpful, may need to be buttressed by direct evidence (e.g., blood tests) under certain circumstances. Similarly, if the statute under which the defendant is being tried defines the crime as operating under the influence of narcotic drugs, the government's case may need to be buttressed by direct evidence (e.g., blood test results). This is because the law here is narrowly drafted to exclude drugs which are not habit-forming. Behavioral tests here are still useful but do not suffice standing alone to obtain a conviction.

A.1.3 Police officers as witnesses. This section deals with the manner of introducing the results of behavioral tests into evidence. A police officer unless shown to be an expert with drugs, may not give his opinion as to whether or not the defendant was under the influence of drugs. He may only describe the defendant's behavior and attitude. Further, a police officer is not allowed to testify that a particular defendant "failed" behavioral tests unless he can first be qualified by the prosecutor as having expertise with the test and with "normal" performance. Even if not qualified as an expert in administration and interpretation of results from behavioral tests, the officer will be permitted to describe the defendant's performance on the tests. Consequently, even though the police officer may not be qualified to give an "expert" opinion on the meaning of performance from behavioral tests, the jury may still draw their own conclusions regarding driving under the influence of drugs from the officer's narrative description of the defendant's physical reactions on the tests.

A.1.4 Constitutional rights. The constitutional right of the accused to remain silent and to refuse to give testimony against himself does not extend to evidence which is deemed to be non-testimonial. Behavioral tests have been held by the courts to be non-testimonial in nature. Therefore, they are not barred from evidence by the constitutional privilege against self-incrimination. However, the police may not use undue force in encouraging a defendant to submit to behavioral tests. Moreover, while the defendant need not be offered the opportunity to consult with counsel as a matter of right before submitting to behavioral tests, the courts may be reluctant to admit the results of tests if the facts show that the defendant requested an opportunity to consult with his counsel and the request was denied by the officers administering the tests. If the defendant does not request an opportunity to consult with counsel but simply refuses to submit to the tests, his refusal to submit to the tests may be admissible as evidence of guilt. Finally, most states have "implied consent" statutes which require a defendant believed to be operating a motor vehicle under the influence of alcohol to submit to a blood, breath, or urine test or lose his license. Such "implied consent" statutes have withstood constitutional attack on the theory that driving is a privilege bestowed by the state and not a right. Should such statutes be expanded to include requiring the motorist to take behavioral tests when accused of driving under the influence of drugs, they undoubtedly would also withstand constitutional challenge.

A.1.5 Due process. In cases where the defendant timely demands a blood test (whether or not he consents to take behavioral tests) the due process clause of the Constitution requires that the state allow him to obtain one at his own expense. If the blood test shows that no drugs were in the defendant's system, the behavioral test results are, of course, discredited. However, if the results of the blood test indicate the presence of drugs, then the circumstantial (indirect) evidence of being under the influence of drugs represented by behavioral test results is corroborated by direct evidence and accurate conviction rate is increased.

A.1.6 Indigent defendants. An indigent defendant may have a right to demand that direct evidence (i.e., blood test) be procured in his case at state expense. Failure to assist the indigent defendant to obtain what may be compelling exculpatory direct evidence may cost the government its case.

A.2 Legal Relevance of Field Behavior Tests

Despite arguments to the contrary by defense counsel, courts have found that results of behavioral tests may help juries decide cases and have admitted such results into evidence.

For example, in the case of State v. Arsenault, 115 N.H. 109, 336 A. 2d 244 (1975), defense counsel argued that the opinions of police officers as to intoxication based on field behavioral tests are not competent evidence and, therefore, should not be admitted into evidence for jury consideration because behavioral tests lack scientific basis. Defense counsel argued that behavioral tests have no established correlation to blood alcohol level, have no established and quantifiable method of evaluation, that judgment of performance is purely subjective, and that behavioral tests lack significant reliability and validity. The court in response said:

> Evidence does not have to be infallible to be admissible. If it is of aid to a judge or jury, its deficiencies or weaknesses are a matter of defense which affect the weight of the evidence but does not determine its admissibility.

The Arsenault court recognized that behavioral tests are admissible probative evidence even though they do not possess the scientific reliability or the same degree of certitude attributed to chemical analysis of blood because they "avoid the shortcomings of casual observation" since they call for precise body movements and a greater degree of coordination than routine standing, walking, or talking. The court recognized that behavioral tests afford the police officer a multitude of stimuli to observe and therefore "enhance the basis of reliability of his opinion" as to whether or not the defendant was under the influence of an intoxicating substance.

A.3 Behavioral Tests as Evidence

A.3.1 Difference between circumstantial and direct evidence. "Circumstantial evidence" is defined as evidence of a collateral fact, that is, of a fact other than the fact in issue, from which, either alone or with other collateral facts, the fact in issue may be properly inferred. It is indirectly probative of a principal fact. People v. Bretagna, 298 N.Y. 323, 83 N.E. 2d 537 (1949) and People v. Butts, 21 Misc. 2d 799, 201 N.Y.S. 2d 926 (1960). "Direct evidence," on the other hand, is evidence which is directly probative of one or more of the principal facts; it tends to establish one or more of the principal facts in issue without the intervention of any other fact. People v. Butts, supra.

A.3.2 Behavioral tests are circumstantial evidence. When being under the influence of drugs is the crime at issue, testimony of a witness based on behavioral tests describing the defendant's condition and behavior is direct evidence as to the defendant's appearance and behavior, but only circumstantial or inferential evidence as to being under the influence of drugs. This is so because behavioral tests do not directly prove that drugs were in the system. A person's appearance and actions while performing behavioral tests may be clearly abnormal, yet the abnormality may not be due to drugs. Impairment of speech and locomotion may be due to any number of other factors; for example, sickness or the result of injury to the nervous system or to the brain.

Because behavioral tests are "circumstantial" evidence of the influence of drugs on the individual they are subject to the limitations of this category of evidence. Nonetheless, as the next subsection demonstrates, circumstantial evidence may suffice, standing alone, to obtain a conviction of operating a motor vehicle under the influence of drugs.

A.3.3 Sufficiency of circumstantial evidence. Circumstantial evidence, in spite of its limitations, may have sufficient probative worth to prove guilt of driving a motor vehicle while under the influence of drugs by the required standard of proof: guilt beyond a reasonable doubt.

A scientific chemical test of the defendant's blood sample, although more reliable and direct evidence of having drugs in the system, may not necessarily be required for the government to sustain its burden of proof. By way of analogy, many courts have specifically held that chemical tests are not required to prove that the defendant had alcohol in his system. The presence of alcohol or drugs may be inferred from descriptions of the defendant's behavior and performance. See People v. Casa, 113 Ill. App. 2d 1, 251 N.E. 2d 290 (1969) where the court of appeals sustained the defendant's conviction even though there was no direct evidence; the evidence against Casa was circumstantial or inferential in that it consisted simply of testimony of the police officer's observations of Casa's behavior. There was no direct evidence that alcohol was found within the defendant's system from a blood test analysis, because no scientific test for intoxicants was conducted. For further authority that circumstantial evidence, per se, is sufficient, see People v. Culp, 537 P. 2d 746 (Colo. 1975): Erwing v. State, 300 S. 2d 916 (1974); People v. Bies, 2 Ill. App. 3d 1001, 276 N.E. 2d 364 (1971); and State v. Reyna, 92 Idaho 669, 448 P. 2d 762 (1968).

A.3.4 Requirements for direct evidence. Direct evidence is needed under some circumstances in order for the government to obtain a conviction for driving under the influence of alcohol or drugs. Cases show that when the government relies totally on circumstantial evidence to prove guilt it may not obtain a conviction if from the circumstantial facts the jury can make a reasonable inference of innocence. See People v. Butts, supra. For a conviction based on circumstantial evidence to be proper it must produce a moral certainty of guilt and exclude any reasonable hypothesis of innocence. Commonwealth v. Wood, 261 Mass. 458, 158 N.E. 834 (1927). Also State v. Kulig, 37 Ohio St. 2d 157, 309 N.E. 2d 897 (1974).

In the case of Seely v. State, 471 P. 2d 931 (Okla. 1970) the defendant contended that the behavior testified to by the police was caused by carbon monoxide fumes from a faulty manifold gasket on his automobile. The court of appeals reversed Seely's conviction because the pathological symptoms testified to reasonably might have been due to exposure to carbon monoxide and because the guesswork could have been taken out of the case simply by administering a chemical test to determine the actual blood content. In Seely there was no evidence that the defendant was stopped because his vehicle was being driven in an erratic manner nor was there evidence of any symptoms or behavior typically associated with being under the influence of an intoxicant except "speech and attitude." The impaired speech was explainable by the carbon monoxide; the defendant's antagonistic attitude developed only after he was placed under arrest and the police were in the process of deciding what to do with the defendant's car and with his wife and children who were in the car. The court of appeals felt that the defendant's hostile behavior likely had etiology other than from an intoxicating substance. In reversing the jury verdict, the appeals court specifically stated that "the state failed to prove the charges 'beyond a reasonable doubt' notwithstanding the jury verdict" of guilty. The court in People v. Mundorf, 85 Ill. App. 2d 244, 229 N.E. 2d 313 (1967) held that circumstantial evidence from observing the defendant did not prove guilt beyond a reasonable doubt in light of testimony that the defendant had recently been hospitalized for an ulcer and liver condition which prohibited him from drinking and corroboration testimony from a passenger that the defendant had nothing to drink. In City of Evanston v. Jaman, 88 Ill. App. 2d 441, 232 N.E. 2d 28 (1967) the court held that evidence that the defendant was involved in an accident and hit his head against the steering wheel, breaking the steering wheel and causing his lip, cheek, and nose to bleed, plus evidence that the defendant had taken pills for an asthmatic condition which made him feel woozy was sufficient to explain the police officer's observations. See also People v. Thomas, 34 Ill. App. 3d 578, 340 N.E. 2d 174 (1975).

The cases of Seely v. State, People v. Mundorf, City of Evanston v. Jaman, and People v. Thomas, cited above, suggest that circumstantial evidence, of alcohol or drugs, even if obtained from behavioral tests, is not enough for the government to get a guilty finding if from the facts the jury can draw a reasonable inference of innocence. When the defendant's condition and behavior can by reasonable inference be attributed to shock and injuries from a collision, the crime of driving under the influence of drugs or alcohol is not proved beyond a reasonable doubt. People v. Merna, 233 App. Div. 739, 250 N.Y.S. 351 (1931); People v. Betts, 142 Misc. 240, 254 N.Y.S. 786 (1931) and People v. King, 28 N.Y.S. 2d 460 (1941).

Although the state enhances the probability of obtaining a conviction by coupling blood tests (where available) with results of behavioral tests, it should be noted that the state is not precluded from obtaining a conviction for operating under the influence of drugs in the absence of blood test results merely because defense witnesses testify contrary to witnesses for the state. This is because it is the function of the trier of fact (the jury, or the judge in a jury-waived trial) to determine the credibility of the witnesses and the weight to be accorded their testimony. For example, in People v. Coolidge, 124 Ill. App. 2d 479, 259 N.E. 2d 851 (1970) the defendant claimed that the conviction of driving under the influence of alcohol was not proved beyond a reasonable doubt because, according to him, his actions after the accident were not induced by alcohol, but rather were a nervous reaction to the accident itself. The defendant in Coolidge argued that his conviction could not stand without corroborating scientific tests showing direct evidence of an intoxicant having been present in his system. The appeals court held that it was within the province of the trier of fact not to believe the defendant's testimony that his actions after the accident were attributable to his nervousness rather than to intoxication.

Citing People v. Casa, supra; People v. Armstrong, 16 Ill. App. 2d 365, 148 N.E. 2d 187 (1958) and People v. Buzinsta, 64 Ill. App. 2d 194, 212 N.E. 2d 270 (1965), the Coolidge court held that the circumstantial evidence given by the arresting officer as to intoxication was sufficient to sustain the conviction even though it was contrary to the defendant's testimony and even though there was no direct evidence from scientific tests against the defendant because it is the jury's right to afford little or no weight to the defendant's testimony.

A.3.5 Effects of distinction between "narcotic" and other drugs. Chemical tests may be necessary for the government to obtain a conviction if the state statute specifically makes it a crime to drive under the influence of "narcotic drugs" and does not specifically make it a crime to drive under the influence of "drugs." This is because there is a category of drugs which are not narcotic or habit-forming.

Under a statute which makes it a crime to operate a motor vehicle under the influence of a "narcotic substance" the "state to convict must demonstrate that the drug used by the defendant was a narcotic or habit-forming drug." State v. Siegmeister, 106 N.J. Super. 577, 256 A. 2d 319 (1969). The Siegmeister court stated that:

> It is not an unknown concept for the courts to require that the State prove that a drug used by a defendant is within the statutory categories ... in order to prove guilt ... the State could not just show that the defendant was under the influence of any drug. Instead, the State must prove that the drug influencing the behavior of the defendant was a narcotic or habit-forming drug.

See also, State v. Tiernan, supra.

If the state statute makes it a crime to operate a motor vehicle under the influence of "narcotic drugs" it seems that circumstantial evidence from behavioral tests alone will not suffice to obtain conviction because there must be a reasonable inference that the drug the defendant was under the influence of is not a non-narcotic drug. Direct evidence from a blood analysis would be needed to obtain this fact and thus a conviction.

It should be noted that the blood test is employed under these circumstances to establish the presence of a narcotic drug in the system. No more need be demanded from the blood test. It need not be used to determine if the defendant is under the influence. The performance test can be used effectively for the purpose of determining whether or not the defendant is under the influence of the narcotic substance.

A.4 Police Officers as Witnesses

A.4.1 Police officers as non-experts. Regarding testimony on the effect of alcohol on the individual, it has long been held that even a non-police officer lay witness may testify that a person was drunk or under the influence of intoxicating liquor without violating the general rule requiring opinion testimony to be given by an expert. See *Pierce v. State*, 173 Neb. 319, 113 N.W. 2d 333 (1962) and *State v. Willard*, 241 N.C. 259, 84 S.E. 2d 899 (1954). Courts have held that a lay person is qualified to give testimony as to intoxication from alcohol because alcoholism is so common and because the law permits a person "experienced in living and everyday affairs" to express an opinion without having any special skill. *People v. Butts*, *supra*. But whereas lay witness testimony standing alone may be sufficient to sustain a conviction for driving under the influence of alcohol, lay witness testimony appears to be insufficient to sustain a conviction for driving while under the influence of drugs. The court in *State v. Tiernan*, 123 N.J. Super. 322, 302 A. 2d 561 (1973) held that a police officer who has no more than a brief one-week training course is not qualified to render an opinion as to whether or not the defendant was under the influence of drugs. In *Smithhart v. State*, 503 S.W. 2d 283 (Tex. Crim. 1974) the court specifically stated:

> Unlike alcoholic intoxication, which is "of such common occurrence" that its recognition requires no expertise ... this court is unable to say that such is the case with being under the influence of drugs.

Tiernan and Smithhart do not prohibit the non-expert police officer from describing his observations of the defendant's performance on behavioral tests. They do, however, prohibit him from giving his opinion to the jury on the ultimate question, i.e., was the defendant operating a motor vehicle under the influence of drugs? A police officer, through training and experience with behavioral tests, can acquire the skills necessary to become an expert witness and thereby testify whether or not in his opinion a defendant was under the influence of drugs. He would have to demonstrate to the court's satisfaction that he is experienced sufficiently with behavioral tests to determine whether or not the defendant had drugs in his system. One obvious way for a police officer to prove his expertise is to show his opinion based on behavioral test results has significantly correlated with the results of blood tests on a broad sample of subjects.

A.4.2 Constraints on testimony. Without special training in the use of behavioral tests, a police officer may not testify that the defendant "failed" the tests unless he is qualified as having expertise in administering the test and knowledge of the test's reliability and validity as well as knowledge of what constitutes "normal" performance on the test. However, it is entirely proper for the police officer to testify what tests or maneuvers the defendant undertook and to recount his observations of the defendant's physical reactions in executing the tests. See, State v. Morton, 74 N.J. Super. 528, 181 A. 2d 785 (1962); aff'd., 39 N.J. 512, 189 A. 2d 216 (1963). Morton stands for the proposition that a police officer who has a motorist perform physical tests may not testify that the motorist failed to pass the tests unless the officer is qualified as an expert witness having special training in the use of the tests. Unless so qualified, the officer may only describe the objective physical responses of the motorist while performing the various tests. The officer may not give a conclusory opinion as to whether or not the motorist's reactions to behavioral tests were a departure from the normal or standard. The Morton court held that such a conclusion may not be given unless the police officer is first shown to have knowledge sufficient to be qualified as an expert to make interpretations from test performance.

The police officer's non-expertise need not present a problem. He need not have to testify whether or not the defendant failed on a behavioral test, if the test is simple and easily communicates to the jury performance and/or judgment skills necessary for safely driving a motor vehicle. The behavioral tests should be non-esoteric if the police officer cannot qualify as an expert with the tests.

A.5 Constitutional Rights and Behavioral Tests

A.5.1 Right to remain silent. The Fifth Amendment to the United States Constitution provides that "No person ... shall be compelled in any criminal case to be a witness against himself." In Malloy v. Hogan, 378 U.S. 1, 84 S. Ct. 1489, 12 L. Ed. 653 (1964) the United States Supreme Court held that the Fourteenth Amendment secures against state invasion the same privileges that the Fifth Amendment guarantees against federal infrirgement. The privilege is a bar against compelling "communcation" or "testimony." The privilege of right to remain silent is not violated by having the accused give evidence non-testimonial or communicative in nature. The Fifth Amendment privilege against self-incrimination does not prohibit the use of the accused's body or aspects of his body such as a blood sample because such evidence is non-testimonial. The Fifth Amendment privilege of right to remain silent is not violated by subjecting the accused to the withdrawal of blood and the admission in evidence of the result of its analysis (Schmerber v. California, 384 U.S. 757, 86 S. Ct. 1826, 16 L. Ed. 2d 908 (1966)); or requiring the accused to don particular items of clothing or speak the words of the perpetrator of a crime (United States v. Wade, 388 U.S. 218, 87 S. Ct. 1926, 18 L. Ed. 2d 1149 (1967)); or requiring the accused to submit a handwriting examplar (Gilbert v. California, 388 U.S. 263, 87 S. Ct. 1951, 18 L. Ed. 2d 1178 (1967)); or voice examplar (United States v. Dionesio, 410 U.S. 1, 93 S. Ct. 764, 35 L. Ed. 2d 67 (1973)).

It can be argued that behavioral tests are "testimonial" or "communicative" in nature and are distinguishable from a blood test, wearing certain clothing, speaking the words of the perpetrator of the crime, or giving a voice or handwriting examplar because behavioral tests require a much greater degree of participation and because there is not the same inevitable result despite the degree of motivation of the defendant and also because with behavioral tests the accused is actively trying to perform acceptably to demonstrate, i.e., "communicate," innocence. See People v. McLaren, 55 Misc. 2d 676, 285 N.Y.S. 2d 991 (Dist. Ct. Nassau County 1967). In spite of the above argument, the courts overwhelmingly have held physical behavioral tests to be non-testimonial and therefore outside the scope of Fifth Amendment protection. Whalen v. Municipal Court of Alhambra, 274 Cal. App. 2d 809, 79 Cal Rptr. 523 (1969) held that field behavioral tests, such as walking heel-to-toe on an imaginary line, finger-to-nose test, and several balance exercises which were administered to the defendant were not violative of his right against self-incrimination. See, Commonwealth v. Kloch, 230 Pa. Super. 563, 327 A. 2d 375 (1974); People v. Killian, 42 Ill. App. 3d 596, 1 Ill. Dec. 297, 356 N.E. 2d 423 (1976) and "Requiring Submission to Physical Examination or Test as Violative of Constitutional Rights," 25 A.L.R. 2d 1407. Also see State v. Arsenault, supra. In Arsenault the court specifically held that performance tests:

> do not seek to compel from the defendant any knowledge he might have, ... nor do they involve the defendant's communicative faculties in any way ... They only compel him to exhibit his physical characteristics of coordination ... Thus the field ... tests do not constitute testimonial compulsion but are merely a source of real or physical evidence and are not within the privilege against self-incrimination. Consequently, the rule of Miranda v. Arizona, 384 U.S. 436, 86 S. Ct. 1602, 16 L. Ed. 2d 694 (1966) is not involved.

A.5.2 Use of force. Although behavioral tests may be compelled, excessive force may not be employed. The Due Process Clause of the Fourteenth Amendment requires that law enforcement officials respect certain decencies of civilized conduct. Evidence obtained in violation of that standard, by means that offend a sense of civility will be excluded from jury consideration. See Rochin v. California, 342 U.S. 165, 72 S. Ct. 205, 96 L. Ed. 183 (1952).

A.5.3 Behavioral tests without counsel. Because the Fifth Amendment right to remain silent has no applicability to behavioral tests in that responses on the tests are non-testimonial, the defendant need not be offered the opportunity to consult with an attorney before undertaking the tests. City of Highland Park v. Block, 48 Ill. App. 3d 241, 6 Ill. Dec. 285, 362 N.E. 2d 1107 (1977). However, if the defendant under arrest asks to consult with counsel prior to making a decision whether or not to submit to the behavioral tests the opportunity should be afforded. State v. Wood, 576 P. 2d 1181 (Okla. 1978).

Although there is no Fifth Amendment mandate that the police permit the defendant to talk with his attorney prior to deciding whether or not to take the behavioral tests, it seems likely that the court would be reluctant to allow police officers to describe the defendant's "poor" performance on the tests if the defendant had been denied access to counsel's advice prior to taking the behavioral tests. Courts might feel that due process requires that defendant's request be honored. Moreover, refusal to take the behavioral tests might not be admissible against the defendant if the defendant refused to submit to the test because his request to consult with counsel was denied.

A.5.4 Refusal to submit to behavioral test may be used against defendant. If the defendant does not ask to talk to counsel and simply refuses to take the behavioral tests the refusal could, at trial, be admissible as evidence against the defendant. This is because behavioral tests are not testimonial and therefore have not been deemed to require the Fifth Amendment protection of right to remain silent.

A.5.5 Implied consent and behavioral tests. Most states have statutes similar to the following Ohio statute regarding alcohol:

> Any person who operates a motor vehicle upon the public highways in this state shall be deemed to have given consent to a chemical test or tests of his blood, breath, or urine for the purpose of determining the alcoholic content of his blood if arrested for the offense of driving while under the influence of alcohol. Ohio Revised Code 4511.191A

Under the "implied consent" statutes, if a defendant refuses to submit to "a chemical test, or tests of his blood, breath or urine" the state may revoke the defendant's license. The courts have upheld the constitutionality of these statutes. Consequently, if the states adopt implied consent statutes regarding behavioral tests, those statutes should also withstand challenge.

A.6 Blood Tests in Addition to Behavioral Tests

The court in Application of Newbern, 175 Cal. App. 2d 862, 1 Cal. Rptr. 80 (1959) held that:

> The Constitution gives to every person accused of a crime ... the right to a fair trial, the right to summon witnesses in his own defense, the aid of counsel, and due process of the law ... The spirit and the purpose of these constitutional guarantees is to assure to everyone a full and ample opportunity to be heard before he can be deprived of his liberty or his property.

In addition to the above, the Newbern court stated that "while it is primarily the function of the courts to see that these rights are not denied, law enforcement agencies also have a responsibility to protect as well as to prosecute," and "that evidence pointing to ... innocence should not be suppressed" because "for an innocent man to be convicted in unthinkable."

For the above reasons, Newbern held that a significant and crucial opportunity to defend against the charge of operating a motor vehicle under the influence is denied a defendant if the police refuse to permit the defendant to call a physician at this own expense for the purpose of taking a blood sample for analysis of intoxicants. The court said that to prohibit the defendant's request is "unreasonable and a denial of due process." The court recognized that "officers charged with law enforcement must always be mindful that the public has as great an interest in the vindication of the innocent as it does in the punishment of the guilty." The court declared that the refusal to allow the defendant a reasonable opportunity to procure possible exculpatory evidence that might establish innocence is as much a suppression of evidence as if existing exculpatory evidence were actually withheld from the defendant.

Although courts hold it to be a denial of Fourteenth Amendment due process of law for the police to prevent a defendant from having a physician at defendant's own expense take a sample of blood, the courts hold that the burden is on the accused to act affirmatively and request that the blood sample be taken. The police are not required to take the initiative even though a blood test is direct evidence and the most compelling evidence of guilt or innocence and, therefore, possibly very useful to the defense of the accused. The case of State v. Reyna, 92 Idaho 669, 448 P. 2d 762 (1968) specifically held that although "the State may not suppress evidence (by not affording the defendant a reasonable opportunity at own expense to obtain a blood sample) ... it need not gather evidence for the accused." The Reyna court further said that the state cannot be said "to suppress evidence simply by omitting to gather evidence unnecessary to its case." For further authority see People v. Culp, supra; People v. Mankowski, 28 Ill. App. 3d 641, 329 N.E. 2d 266 (1975); State v. Urrego, 41 Ohio App. 2d 124, 322 N.E. 2d 688 (1974) and Hendrix v. State, 125 Ga. App. 327, 187 S.E. 2d 557 (1972). The state, according to People v. Culp, supra, has no duty under the Due Process Clause of the Fourteenth Amendment to obtain for the defendant what might be exculpatory evidence.

In spite of the above authority, an argument can be made that due process of law is an evolving concept and should require that the state take the initiative and offer the defendant a blood test.

Brady v. Maryland, 373 U.S. 83 (1963) began the trend toward expanding rights of the accused regarding preserving and disclosing exculpatory evidence. In Brady the United States Supreme Court held that:

> Suppression by the prosecution of evidence favorable to an accused upon request violates due process where the evidence is material ... to guilt ... irrespective of the good faith or bad faith of the prosecution.

United States v. Heath, 147 F. Supp. 877 (Hawaii, 1957), aff'd, 260 F. 2d 623 (9th Cir. 1958) and Trimble v. State, 75 N.M. 183, 402 P. 2d 162 (1965) both hold that a duty to disclose exculpatory evidence may necessitate affirmative acts by the government. Also see, Note, Government Has Duty to Implement Effective Guidelines to Preserve Discoverable Evidence, 1971 Duke L.J. 644. Prosecution argument that the duty to preserve blood sample does not exist if not timely demanded by the defendant appears to be inconsistent with the above authority and with United States v. Bryant, 439 F. 2d 642 (D.C. Cir. 1971) where the court stated "before a request for discovery has been made the duty of disclosure is operative as a duty of preservation." The Bryant court looked to Brady v. Maryland, supra, and held that since "misplaced" tape recordings used during the government's investigation "might" have been favorable to the defendant, the prosecution had a duty to disclose them. In the case of People v. Johnson, 3 Ill. 2d 602, 203 N.E. 2d 399 (1966) the court said regarding grand jury testimony that there is no reason not to preserve it unless the state is interested in running the risk of convicting accused parties on the testimony of untrustworthy persons. The Johnson court said that if a defendant is precluded at trial from impeaching a witness through the witness's inconsistent testimony given in front of the grand jury justice would be defeated. By analogy, it can be argued that there is no reason why the government should not take the initiative and offer the defendant a timely blood sample if the government wants to minimize the risk of convicting an innocent person. Regarding grand jury testimony, the United States Supreme Court held in Dennis v. United States, 384 U.S. 855, 874 (1966) that courts may not assume the absence of inconsistencies between trial and grand jury testimony. Likewise, it seems that it should not be assumed that a blood analysis will not exonerate a defendant charged with operating under the influence of drugs.

Certainly, the better practice would be to notify the defendant of the right to a blood test at his own expense by his own physician or by a lab technician. People v. Ward, 307 N.Y. 73, 120 N.E. 2d 211 (1954). Doing so would meet and answer the defense if no drugs are found to be present. Also, the state would be supplied with direct evidence of drugs if the defendant takes a blood test that turns out to be positive. This would relate aberrant performance on behavioral tests to drugs and thereby serve to increase the conviction rate. Again, it should be noted that in states where the criminal statute requires that the drug must be a narcotic drug, to prove the defendant to have been in violation of the statute of operating a motor vehicle while under the influence of drugs may, of necessity, entail the offering of a blood test. So, for this reason also, the better practice would be to offer the defendant a blood test.

A.7 Rights of Indigents

Because a blood sample showing a person to be free of narcotic substances is the most compelling evidence (more compelling than behavioral tests) and because a non-indigent defendant has a right under the due process clause of the Fourteenth Amendment to demand a timely blood test by his own physician at his own expense, it might be argued that an indigent defendant who timely requests a blood test at government expense must be afforded such a test under the equal protection clause of the Fourteenth Amendment at the state's expense. In *Commonwealth v. Tessier*, 360 N.E. 2d 304 (Mass. 1977) the court held that a statute providing that a defendant charged with operating a motor vehicle while under the influence of intoxicating liquor has a right at own expense to be examined by a physician personally selected is not violative of the equal protection clause as applied to an indigent defendant even though the statute fails to provide for an independent examination at public expense for the indigent defendant. The court in so holding reasoned that an indigent defendant is afforded the opportunity to be tested by an apparently qualified police breathalyzer operator and that such testing satisfies constitutional standards. Applying the reasoning of *Commonwealth v. Tessier*, *supra*, to testing in the area of driving under the influence of drugs, one might plausibly argue that if the police offer the indigent defendant only behavioral tests, the defendant may under the Fourteenth Amendment Equal Protection of Law Clause be entitled, upon demand, to a blood test at state expense. This may be so because a blood sample free of drugs would be very important evidence and because

Fourteenth Amendment Equal Protection requires that an indigent defendant not be deprived of a viable defense because of indigency.

Tessier effectively supports this theory since that case appears to have turned on the availability of a reliable alternative to the physician's clinical test whereas in the area of drug influence on behavior no such alternative exists. Where no reliable alternative to blood sampling is available, the government may be required to furnish a test to an indigent defendant or risk losing its case.

A.8 Conclusion

In general, it is clear that behavioral tests have an important role to play in the building of a case by the prosecution to obtain a conviction for the crime of driving under the influence of drugs. It has been shown that although behavioral tests are circumstantial evidence of the influence of drugs and although they cannot be employed to distinguish between the presence in the system of narcotic drugs as opposed to non-narcotic drugs, they can nonetheless provide important evidence. Behavioral tests can show whether or not a defendant's motor coordination and/or judgment are impaired.

By contrast, blood tests, while direct evidence of the presence of a drug in the system, cannot provide evidence of impaired behavior unless the state is willing to go to great expense. This is because the scientist who is analyzing the blood sample would have to determine and testify as to what quantity of a particular drug would affect a person with the defendant's tolerance, constitution, and weight to a sufficient extent that he would be deemed or presumed to be under the influence of the drug to the extent required by the statute. Because there are so many drugs or combinations of drugs which might be used by a defendant, the scientist's analysis, to be accurate, would have to be complex and time-consuming. The cost of such analysis would be prohibitive to most police departments.

The most effective and cost-efficient use of testing seems to be the combining of a simple blood analysis with behavioral tests. A simple blood analysis which tests only for the presence or absence of drugs and determines whether or not the drug is narcotic or non-narcotic if coupled with behavioral tests which illustrate impaired behavior, is much less expensive than testimony from a chemist or physician regarding the effects on a defendant's behavior of minute percentages of a drug or drugs in the blood stream. Moreover, the results of simple blood analysis and performance on behavioral tests are much easier for the trier of fact to understand than the complex and esoteric inferences derived from a detailed blood analysis.

This research has shown that a combination of direct evidence from a blood test and circumstantial evidence from behavioral tests not only enhances the probability of the government's obtaining a conviction but also safeguards the government's case from constitutional challenge because by conducting a blood sample test the government has gathered evidence potentially helpful to the defendant. The defendant cannot argue that he was deprived of potentially exculpatory evidence.

Finally, the prosecution is most likely to obtain a conviction if it keeps the evidence as straightforward and as simple as possible. The prosecutor must not forget that the audience, the jury, is composed of lay persons who will not understand the relationship between behavioral test results and the defendant's ability to operate a motor vehicle safely unless the relevance of the test is readily apparent. If esoteric tests are used or if the behavior elicited from the tests does not easily relate to driving, the police departments might have to incur the cost of employing experimental psychologists who will have to educate the jury on how the skills demanded on a particular behavioral test relate to the skills required to operate a motor vehicle. Needless to say, such expense and confusion should be avoided. Expert witnesses will not be needed if behavioral tests are simple and require performance skills that a jury can readily associate with operating a motor vehicle.