American Association of
Motor Vehicle Administrators

Licensing
Rules
of the Road
MODEL TESTING
Requirements

# AAMVA GUIDELINES for Noncommercial Knowledge and Skills Test Development

Supports the Noncommercial
Model Driver Testing System

*Version: September 2014*

# PREFACE

These guidelines were originally prepared by A. James McKnight on behalf of the Driver License Committee "Driver Testing Working Group" of the American Association of Motor Vehicle Administrators in 1997.

The guidelines were last revised in September 2014 by Highway Safety Services, LLC to support the development and maintenance of the 2007 Noncommercial Model Driver Testing System (NMDTS) and to provide assistance to AAMVA and the Test Maintenance Subcommittee (TMS) in working with other organizations concerning proposals for additional language for the model driver manual and knowledge test item pool.

Additionally, these guidelines were developed to assist jurisdictions with the development and evaluation of their own test questions to support jurisdictional specific content and laws.

These guidelines served as the basis of the NMDTS developed by the AAMVA TMS and Highway Safety Services, LLC.

The NMDTS materials were pilot tested with cooperation from the Indiana University of Pennsylvania (IUP) – Highway Safety Center and Field Tested in the states of South Carolina and Maryland (see NMDTS Final Report to AAMVA/NHTSA). Results of the Pilot and Field Studies were applied to the final version of the NMDTS materials.

# TABLE OF CONTENTS

# INTRODUCTION

This report provides guidance for the development and evaluation of tests to assess the knowledge and skills of applicants for a license to operate automobiles. It has been developed by the American Association of Motor Vehicle Administrators (AAMVA) as part of an attempt to help State licensing agencies achieve uniformly high quality in assessing the ability of driver license applicants to operate vehicles in a manner that assures the safety and mobility of the driving public.

The development and evaluation of driver license tests in the past has often been hampered by the inappropriate application of psychometric concepts and techniques. Psychometric testing is largely intended to measure constructs defined by tests themselves and validated in terms of their ability to predict future behavior. Driver license tests, on the other hand, attempt to measure knowledge and skills defined by motor vehicle agencies and are intended to serve as an incentive to applicants to acquire the skills, knowledge and attitudes needed for safe driving and as a means of making sure applicants possess the skills and knowledge needed before they are issued a license. The difference in the purpose of tests leads to substantial differences in the way the license tests and psychometric measures are developed and evaluated, differences that will be noted throughout the report.

The scope of the guidelines is limited to the initial licensing of drivers in general. It does not address the unique requirements of licensing for special vehicles such as trucks, buses, or motorcycles. Nor, does it attempt to accommodate drivers with specialized needs, such as drivers with disabilities, the aging/mature driver, frequent traffic offenders, or drinking drivers. Within the general driving population, the guidelines do address those whose language limitations interfere with their ability to acquire, or to demonstrate the possession of, the driving knowledge for which they are held responsible.

The body of these guidelines is divided into two sections, corresponding to the two types of abilities to be tested: knowledge and skill. The guidelines for *knowledge testing* cover both the written test and the driver manual from which test questions are drawn. The manual and test represent two essential elements of the knowledge testing process. Recommendations are also provided for administration of tests through automated processes and the special needs of testing drivers with limited reading ability. Guidelines for *skill testing* focus upon road tests, as they are the most common means of assessing driving skills within the licensing structure. However, guidelines are also provided for vehicle safety inspection testing, off-street testing and testing with the aid of simulation.

# GUIDELINES FOR KNOWLEDGE TESTING

This section of the guidelines describes methods for assessing the knowledge of driver license applicants. Research has shown that a license testing program directed at critical knowledge requirements is capable of reducing the likelihood that drivers would be involved in accidents for which they are responsible.

The purpose in giving knowledge tests is to assure that driver license applicants possess the information required to operate vehicles in a way that is consistent with the safety and mobility of the public. Providing this assurance means not only assessing applicant knowledge through a written test, but providing a manual or means by which applicants can acquire that knowledge. The two elements of knowledge testing are equally important. These guidelines will address *knowledge requirements*, the *driver manual*, and the *written test*, as well as the special requirements of applicants with language limitations. The scope of the guidelines is confined to testing the knowledge of applicants for a Basic Driver's License. Other guidelines will deal with requirements of heavy vehicles and motorcycles, as well as the special needs of older drivers and drivers with disabilities.

## KNOWLEDGE REQUIREMENTS

The development of both the driver manual and the written test derive from a common source — the definition of the knowledge required to enable drivers to operate their vehicles in a way that is consistent with the safety and mobility of the motoring public.

At one time, the content of driver manuals and knowledge tests was confined to laws and regulations governing motor vehicle operation. This restriction reflected the position that drivers could only be held accountable for knowing what was imposed upon them by law. However, it is now generally accepted that applicants can be held responsible for any knowledge that contributes to the safety and mobility of the public, so long as the necessary information is made available to them through the driver manual or some alternate source. Knowledge requirements include, in addition to laws and regulations, driving procedures, principles, facts, and concepts, including both those that *enable* drivers to operate their vehicles properly and those that *motivate* them to do so. A list of knowledge categories appears in the table on the following page. An initial set of knowledge requirements had been identified through earlier research involving a comprehensive analysis of driver tasks and their prioritization in terms of their criticality to traffic safety. This set of knowledge requirements was disseminated among State and Provisional license agencies for review and suggest additions, deletions, and revisions.

## TABLE 1

## KNOWLEDGE/SKILLS REQUIREMENTS

### PRE/POST DRIVING

**Trip Planning**

**Adjustments**
- Seat Position
- Mirrors

**Occupant Protection**
- Safety belts
- Air bags
- Locked doors

**Inspection**
- Vehicle walk around
- Leaks
- Tires
- Lights
- Turn signals
- Windows and windshield
- Wipers and washers
- Heater and defroster
- Horn
- Indicator lights
- Loose objects
- Braking system
- Steering system
- Suspension system
- Exhaust system
- Engine

**Cleaning**
- Windshields and windows
- Mirrors
- Lights

**Securing Vehicle**

### VEHICLE CONTROL

**Starting**
- Starting procedure
- Limited warm-up

**Accelerating**
- On the flat
- On upgrades
- On slippery surfaces

**Shifting (Manual Transmission)**
- Shift at proper speed/rpm
- Coordinating clutch/acceleration

**Steering**
- Hand position
- Hand over handHand to hand

**Staying in Lane**
- Grasping wheel
- Adjusting wheel to turn to speed and position
- Fixate well ahead

**Turning**
- Positioning for turn
- Adjusting speed for turn
- Turning wheel in relation to speed and path
- Straightening wheel

**Regulating Speed**
- Regulating accelerator to maintain speed
- Observing speedometer
- Keeping transmission in gear

**Slowing/stopping**
- Anticipating stops
- Applying brake
- Easing brake at stop
- Maintaining brake pressure when stopped
- Stopping distance

**Special Handling Characteristics**

**Backing**
- Assuming proper body position
- Observing through rear window
- Coordinating clutch and accelerator
- Turning wheel in relation to speed and path
- Braking to a stop

### RULES OF THE ROAD

**Traffic Controls**
- Traffic lights
- Stop signs
- Yield signs
- No-turn signs
- No enter signs
- Crosswalks
- Railroad crossing signs/lights
- Human controls enforcement/highway personnel)
- Work zone signs
- Guide signs
- Route number signs

**Lane Control**
- Basic lane use
- Passing
- Reversible lanes
- Reserved lanes (eg, HOV)
- Shared left-turn lanes
- Backing
- Stopping
- One-way
- Lane drops, merges

**Turns**
- General rules
- Turn control signs
- Turnabouts
- U-turns
- 3 Point turns
- Roundabouts and traffic circles

**Right-of-Way**
- Yielding right-of-way
- Intersections
- Roundabouts and traffic circles
- Pedestrians
- Emergency vehicles
- School buses

**Vehicle restrictions**

**Parking Restrictions**

### VISUAL SEARCH

**Maintaining Attention**
- Maintaining general surveillance
- Avoiding distraction

**Search Ahead**
- Distance
- Side-to-side

**To the Side**
- Intersections
- Crosswalks
- Railroad crossings
- Roadside activity
- Sight obstructions
- Merges/on-ramps

**Over-the-Shoulder**
- Lane change
- Merging

**Mirrors**
- Periodic scanning
- When slowing

Changing lanes
Merging
Overtaken on downgrades

**Headlight Use**
Use of high beams
Dimming for vehicles
Low beams for fog and rain
Not retaliating
Night driving

## COMMUNICATION

**Signaling Intentions**
Signaling turns
Nature
Timing
Canceling signal
Signaling slow/stop
Uses hand signals when appropriate

**Communicating Presence**
Headlights
Horn
Emergency flashers
Signals (reflectors, flares)

## ADJUSTING SPEED

**Compliance with Limits**

**Adjusting to Traction**
Slick surfaces
Curves
Hydroplaning

**Adjusting to Visibility**
Intersections
Hills, curves
Vehicles
Weather
Darkness
Fog

**Adjusting to Traffic**
Prevailing speed
Entering traffic
Leaving traffic
Pulls over when required
Emergency vehicles

**Specific Hazards**
Maneuver limitations
Roadside activity
Path threats
Pedestrian traffic
Shopping areas
Wildlife

## POSITIONING VEHICLE

**When Following**
Vehicles in general
Specific vehicles
Limited visibility
Avoiding blind spot
Slippery surfaces
When carrying/towing heavy loads
When followed

**Passing Vehicles**
Gap acceptance (2-3 lane)
Lateral separation

**Crossing/entering**
Accepting proper gap
Assuring clearance ahead
Responding to turn signals
Vision obstructed

**When stopping/parking**
Selecting locations
Vehicle orientation
Keeping clearance
Observes restrictions

## HANDLING EMERGENCIES

**Vehicle Failures**
Brake
Tire
Power
Accelerator
Headlight
Car Fire
Engine Overheating
Carbon Monoxide Poisoning

**Collision Avoidance**
Quick stop
Manual and ABS
Quick turns
Skid recovery
Escape paths
Pavement shoulder drop offs
Rollovers

**Accident procedures**
Scene control
First aid
Summoning help

## SHARING THE ROAD

Pedestrians
Bicyclists
Motorcycles and mopeds
Emergency vehicles
Commercial vehicles
Public transportation
Funeral processions
Slow moving vehicles
Work zones

## SPECIAL DRIVING SITUATIONS

Rural road driving
Night driving
Driving on flooded roadways
Vehicle submerged underwater
Winter driving
Mountain driving
Desert driving
Driving in very hot weather
Avoiding collisions with animals

## DRIVER PREPARATION

**Physical Fitness**
Vision checks
Hearing checks
General physical checks
Treatment for illness/disability
Eating
General
During trips
Exercise
Feeling of motion
Fatigue prevention

**Use of Alcohol and Other Drugs**
Limiting consumption
Limit of driving
Avoiding mixing

**Driver Distractions**
Cell phones
Adjusting radio, CD or climate controls
Adjusting or using GPS or navigation systems
DVD players
Dashboard control panel
Grooming
Talking to passengers
Eating, drinking or smoking
Reading
Picking up something that fell
Outside traffic/vehicle
Police pulling someone over
Sunlight/sunset
People/objects in roadway
Crash scene
Road construction
Reading billboards

**Road Rage**
- Definition
- Signs of road rage
- What to do when a driver has road rage

**Aggressive Driving**
- Definition
- Increase of aggressive driving
- Prevention
- What to do when a driver is aggressive

## VEHICLE READINESS

**Characteristics**
- Vehicle size
- Engine size

**Drive train configuration**
- Displays (legibility)
- Controls (ease of reach, operation)
- Seats
- Trailers and towing

**Safety Equipment**
- Passive restraints
- Mirrors
- Anti-lock brakes
- CB radio

**Inspection/Maintenance Servicing**

## LICENSING

- Types of licenses
- Organ donor

## HOW TO PREPARE FOR YOUR DRIVING TEST

- Pre-Trip vehicle safety inspection test
- Basic vehicle control skills test
- Road test

## DRIVER MANUAL

Licensing authorities in all states and provinces need to provide applicants with a written source of the information required to meet the knowledge requirements specified in the preceding section. The driver manual defines what it is that drivers are expected to know and will be held responsible for knowing. These guidelines will address driver manual content, organization, and format.

*Content* — The subject matter of the driver manual should encompass, at the minimum, all of the knowledge requirements specified in the table above. In addition to content intended to benefit the safety and mobility of road users, other information may be included (e.g., licensing, driver services, organ donor program). Because the driver manual is so widely distributed to and read by the public, it serves as a valuable avenue of communication with the public. However, the inclusion of other topics should not compromise the ability of the manual to serve its primary function.

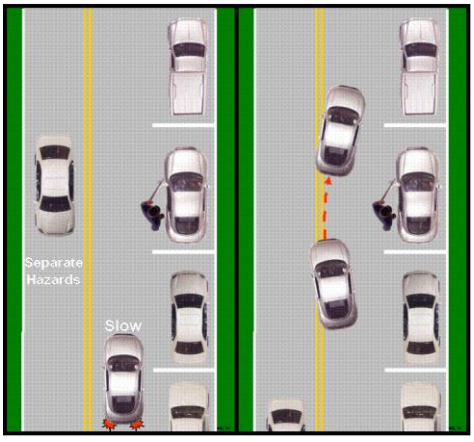
*Organization* — The driver manual is intended primarily to serve as a reference aid and should be organized in a manner that will facilitate access to the individual items of information on an as-needed basis. Such facilitation is achieved by use of: (1) relatively brief, self-contained sections, (2) headings that clearly identify the content of each section, and (3) a detailed subject index.

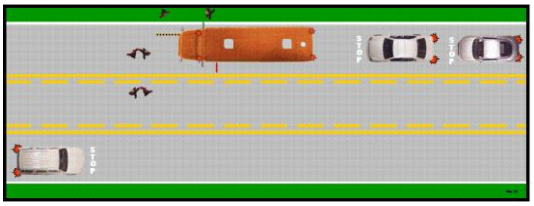
*Format* — The format of the driver manual should be designed to foster acquisition and retention of information. For doing so, it should make liberal use of:

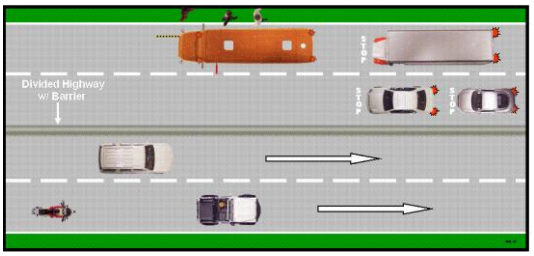
- Short paragraphs
- Bulleted phrases
- Paragraph headings
- Highlighting of key words
- Practice questions

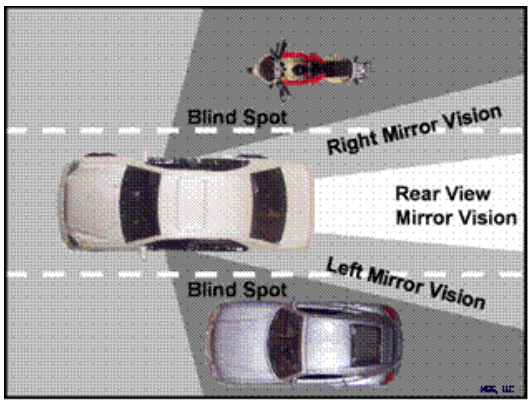
*Reading Level* — The information underlying proper vehicle operation is generally simple enough to be communicated at relatively low reading levels. The fifth- to sixth-grade reading level is considered optimum in communicating with the population of literate driver license applicants. The reading level of the entire manual should be checked before it is published. Available automated procedures for measuring reading levels may be applied quickly and inexpensively to computerized text.

*Visuals* — Pictures, diagrams, and other graphic displays frequently communicate better and give a more lasting impression than text.
However, they tend to be expensive of space and should only be employed where they provide a clear benefit. Their use for clearly ornamental purposes is wasteful and can be deleterious wherever a limitation on the number of pages would force deletion of important content.








*Practice Questions* — The provision of practice test questions will enable applicants to assess their own knowledge and help those who are unfamiliar with the multiple-choice format to gain some experience in its use. It is best that practice questions not be drawn from the specific items making up the knowledge license test. To pass the test, applicants must study the whole manual, not just the practice test questions.

## KNOWLEDGE TESTS

The primary purpose of a knowledge test in driver licensing is to assure the applicant's possession of the information needed to drive safely. It fulfills this purpose by (1) providing an incentive to applicants to secure from the driver manual the information needed to pass the test, and (2) furnishing a means by which applicants can demonstrate their possession of information.

### Test Construction

The construction of knowledge tests will be discussed in terms of (1) content, format and wording of items, (2) scoring standards, and (3) alternate forms.

### Content of Knowledge Tests

The knowledge test can only measure a *sample* of what applicants know. However, if the sample of items is sufficiently large, and represents the full-range knowledge requirements, the test will provide a reliable estimate of what applicants know about the subjects that make up the manual.

The content of test items should be drawn directly from the driver license manual. If the manual has been designed to fulfill the knowledge requirements that underlie safe driving, the test will also be designed to fulfill the requirements. Selecting test content directly from the manual also guarantees that the applicant will have had an opportunity to master it. To the fullest extent possible, the test items should be drawn from across all sections of the manual so that applicants know that any item of information found in the manual may appear in the test. Wording items in the same manner as the text will help examiners demonstrate to applicants that the item did indeed come from the manual.

*Item Format*

The multiple-choice-type of item offers the only practical means for testing large numbers of license applicants uniformly and objectively within the resources generally available to licensing agencies. True/false questions should never be used in a licensing test. The following considerations should guide the design of multiple-choice-type format:

*Content of alternatives* — All of the alternative responses to an individual item should address the same piece of information and attempt to assess whether the applicants possess *that* information. If the various choices address different topics, there is no way of determining from responses what it is that an applicant does and does not know.

*Correct and incorrect answers* — Each item should have only one correct answer; the rest should be clearly incorrect. Applicants should not be required to judge degrees of correctness (which is the "most correct" answer).

*Number of alternatives* — Generally speaking, the greater the number of alternative responses, the smaller the chance of guessing the correct answer. However, the situation applies only where all alternatives are plausible. In driver license exams, it may be difficult to develop more than three alternatives that are plausible. Adding a fourth alternative that nobody chooses makes a test longer without making it better.

*True-False* — The true-false format should be avoided owing to (1) the relatively high probability of guessing the correct answer, and (2) differences in the interpretation of "true" and "false". Knowledgeable applicants are often scored incorrect because they know of exceptions to what are scored on the test as true statements.

*Position of correct answer* — The position of the correct answer in the series of alternatives should be decided by chance in order to prevent applicants from benefiting from systematic patterns, such as a tendency to put the correct answer in the middle of the series.

*Sequence of alternatives* — Where alternatives follow a numerical or otherwise logical sequence, they should appear in that sequence on the test. To preserve the sequence, the alternative responses may have to be selected after the position of the correct alternative has been decided by chance.

**Wording of Items**

Items should be worded to maximize the likelihood that applicants who know the answer will answer the item correctly and those who do not know it will answer incorrectly (validity). In order to achieve this objective, the following should be avoided:

*Complex words or phrases* — The test should measure driving knowledge, not verbal skill.

*"All of The Above"* — In this type of question, all of the alternatives are actually correct. Applicants may read no further than the first alternative. The same hold true for both “x” and “y” are correct.

*"None of The Above"* — In those cases where this is the correct response, there is no way to determine whether an applicant knows what the correct answer truly is.

*Legalese* — What is written by and for lawyers is not necessarily understood by the public. Avoid legal terms and direct excerpts from the motor vehicle code. If the wording is taken directly from the manual, this will not be a problem.

*Use of the negative form* — A question that starts "Which of the following is **not**..." requires applicants to search for an incorrect answer. Knowledgeable applicants frequently forget this and choose the correct answer.

*Inconsistent alternatives* — Inconsistencies that attract attention to a particular alternative should be avoided, examples being alternatives that are substantially longer than others, use of attractive words such as "safely," or including a rationale for incorrect answers to make them appear more plausible.

Licensing authorities should make every effort to prevent applicants from passing the test simply by memorizing the answers to a limited number of test questions. The best means of achieving this objective is by drawing from such a large pool of test items that anything appearing in the driver manual may show up on the test. The availability of a large test item pool permits development of many alternative forms and, with computer testing, generation of a virtually unique test for each applicant.

These practices prevent applicants from gaining high scores simply because they have taken the test before.

**Knowledge Domains**

The larger set of test questions should be drawn from the smaller set of knowledge domains. Test questions are divided into subsets of similar topic areas of knowledge domains. Using the list from Table 1 the following demonstrates the categorization of knowledge areas into a smaller subset of knowledge domains from which questions can be drawn.

**TABLE 2**
**KNOWLEDGE DOMAINS**

**PRE/POST DRIVING**
- Trip Planning
- Adjustments
- Occupant Protection
- Inspection
- Cleaning
- Securing Vehicle

**VEHICLE CONTROL**
- Starting
- Accelerating
- Shifting (Manual Transmission)
- Steering
- Staying in Lane
- Turning
- Regulating Speed
- Slowing/stopping
- Special Handling Characteristics
- Backing

**RULES OF THE ROAD**
- Traffic Controls
- Lane Control
- Turns
- Right-of-Way
- Vehicle restrictions
- Parking Restrictions

**VISUAL SEARCH**
- Maintaining Attention
- Search Ahead
- To the Side
- Over-the-Shoulder
- Mirrors
- Headlight Use

**COMMUNICATION**
- Signaling Intentions
- Communicating Presence

**ADJUSTING SPEED**
- Compliance with Limits
- Adjusting to Traction
- Adjusting to Visibility
- Adjusting to Traffic
- Specific Hazards

**POSITIONING VEHICLE**
- When Following
- Passing Vehicles
- Crossing/entering
- When stopping/parking

**HANDLING EMERGENCIES**
- Vehicle Failures
- Collision Avoidance
- Accident procedures

**SHARING THE ROAD**

**SPECIAL DRIVING SITUATIONS**

**DRIVER PREPARATION**
- Physical Fitness
- Use of Alcohol and Other Drugs
- Driver Distractions
- Road Rage
- Aggressive Driving

**VEHICLE READINESS**
- Characteristics
- Drive train configuration
- Safety Equipment
- Inspection/Maintenance
- Servicing

**LICENSING**

**HOW TO PREPARE FOR YOUR DRIVING TEST**

**Scoring Standards**

Individual States will decide the number or proportion of test items that must be answered correctly in order for the applicant to pass the test. Scoring standards should have relatively little effect upon licensing since almost all applicants will eventually and ultimately pass. What they can influence is the knowledge levels of the licensed population; the higher the standard, the more people will know.

*High Standards* – Fear of burdensome re-testing has encouraged some administrators to accept relatively low scoring standards; the idea that 75% is passing has a long history in education. A number of considerations encourage higher standards, such as 90%, for the driver license test.

- Most of those taking the knowledge test are new drivers, lacking the skill and road savvy that comes with experience. For such a population, a high standard can be justified.

- The content of the test is bounded by the content of a manual that can be mastered in a few hours of preparation. Being informed of the scoring standards in advance will encourage the preparation needed to pass the first time.

- A test is not completely accurate in measuring what applicants know. On a test with a 90% scoring standard, many applicants will pass the test knowing less than 90% of the information from which test items were drawn.

*Differential standards* — Recognizing that certain items of information are more critical to proper operation than others, agencies may wish to set differential standards, requiring close to a 100% for a subset of items in which lack of information poses a clear threat to the public safety e.g., traffic signs and lights or right-of-way laws, and a 90% standard for most other items.

*Improving applicant performance* — Where an unacceptably high proportion of applicants fail to meet established passing standards, efforts should be undertaken to seek improvements in applicant knowledge and/or test procedure rather than lowering the test standards. Such improvements may include the following steps:

- Examining individual items to identify the specific items that are causing trouble,
- Revising the test to clarify any ambiguous questions and eliminate unnecessarily fine distinctions, and
- Revising the treatment of the corresponding subject matter in the driver manual where the test items appear valid, giving it greater visibility and/or improving the effectiveness of communication.

**Alternate Test Forms**

Alternate forms of the knowledge test should be available for administration as retests, thus minimizing the chance of an applicant's being able to answer questions correctly because of previous exposure to the same questions on an earlier test administration. Development of alternate forms should adhere to the following:

*Representative sampling* — Each alternate test form should sample representatively across all knowledge categories or domains, in order to provide the best possible estimate of an applicant's total knowledge.

*Independence of forms* — The questions making up the various alternate test forms should assess different items of information, not simply the same information with different wording. Getting a second chance to answer the same questions does not show what the applicant knows about driving. If it is absolutely necessary to test for the same information on more than one form, at least the incorrect alternatives should be different.

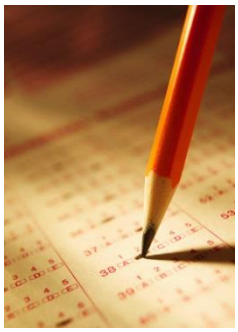
*Equality of forms* — All test forms should have the same level of difficulty, as indicated by the mean proportion of items answered correctly. Equality may be achieved either by allocating individual items to forms on the basis of their difficulty level or through the use of norms that render the forms statistically equal.

*Item independence* — The various items appearing on any one form must be independent of one another. No question should be capable of being answered correctly solely on the basis of information supplied by another item appearing on the same form.

*Re-testing* — Since re-administration of the same test form may yield a spuriously high estimate of applicant knowledge, the chances of such occurring should be minimized by (1) keeping a record of the first to assure use of an alternate form on retest, or (2) having such a large number of alternative forms available that the chances of getting the same form twice are extremely small.

*Computer generated tests* — One means of meeting the requirements just described is by having the entire item pool stored in computers and individual items selected at random for each applicant. While automated testing obviously lends itself to such an approach, computers can print out copies for paper and pencil testing. The possible number of test "forms" would be extremely large. With a large enough item pool and adequate test length, individual forms would be representative and equal in overall difficulty. The selection of items can be programmed to assure these conditions prevail and to prevent the answers to one item appearing in the stem of another question.

**Item Analysis**



As a part of test development, item analyses should be carried out on each question and form of the test as a means of identifying deficient items. Each form should be administered to a representative sample of no less than 100 applicants before and after reading the driver manual. Items should be analyzed for response frequencies and item-test relationships.

*Response frequencies* — The proportion of applicants answering each item correctly should be examined as clues to possible deficiencies in the wording of items.

- Items with extremely low post-test pass rates and those showing little pre-post improvement should be examined to make sure they are not misleading in some way. On the other hand, if the percent choosing the correct answer is close to 100%, wording should be examined to make sure that the correct answer is not being given away.

- The proportion of applicants picking each *distracter* (incorrect alternative) should also be examined. Incorrect alternatives that are never chosen should be examined closely to see if they lack plausibility.

- It is common psychometric practice to seek items with close to a 50% pass rate in order to maximize test variance and potential correlation with other variables. However, while controlling item difficulty is an acceptable practice where the variable being measured is a hypothetical construct defined by the test itself, license test items are drawn from a defined body of content and pass-fail rates are what they turn out to be.

*Item-test relationships* — The relationship between performance on each item and total test score should be examined through the use of item-test statistics. Since all of the questions on the test came from the same source — the driver manual — applicants who do well on the test in general should also do well on each individual item. A weak item-test relationship suggests that something in the wording may be causing knowledgeable applicants to reject what is supposed to be the correct answer, or pick an alternative that also happens to be correct.

The most common measure of item-test relationship is the item test correlation. Two shortcomings of a simple correlation coefficient in this application are (1) the difficulty in determining just how much of a relationship is due to the spurious effect of the item on total score and (2) the inability to detect which alternative is causing the trouble. A more informative technique is to compute the mean total score for the applicants selecting each of the alternative responses, subtracting "1.0" from the mean for the correct alternative to remove the effect of that item on the total score. A distracter having a higher mean score than the correct answer is a potential trouble source.

*Use of item analysis* – Where item analyses identify deficient items, every effort should be made to discover and remedy the deficiency. Wording of the items should be examined and individual alternatives identified and rewritten. Generally, where the mean score of applicants picking a particular incorrect answer is higher than the score for those picking the correct answer, that alternative is likely to be the source of the problem. More often than not, something in the wording of that item is misleading knowledgeable examinees. Corresponding sections of the driver manual should also be scrutinized for messages that might be unclear. However, no item should be deleted from a criterion-referenced measure such as a license test purely on the basis of item analysis results; the content of the knowledge test should be based upon what drivers need to know, not item statistics. Not every piece of information lends itself to the multiple choice format and some items will prove unsalvageable. However, exclusion should not be based upon the results of item analysis alone.

**Test Administration**

The way in which knowledge tests are administered can strongly influence their effectiveness in assuring the safety of drivers.

**Feedback to Applicants**

Time permitting, applicants can be apprised of their errors, informed of the correct answer and told where the information may be found in the driver manual. Where applicants have failed a test, this form of feedback may enhance the credibility of the test and prevent protests.

While providing correct answers also serves as a learning function, the benefits are relatively small since the questions with which applicants are confronted on any one test constitute a small portion of the total information for which they are held responsible. While applicants should be informed of their errors, they should be advised that these errors are indicative of wider information deficiencies that can only be remedied by studying the entire driver manual.

**Scheduling Re-tests**

Applicants who clearly fail the knowledge test should be required to wait at least a day before being re-tested. Some applicants may seek an immediate retest in the mistaken belief that they can pass simply by looking up answers to the questions they missed, an unlikely event where alternate forms are administered. Requiring a day's wait provides applicants an opportunity to restudy the manual in its entirety and thus prepare for any test form they might receive.

**Test Security**

Licensing authorities should make every effort to prevent copies of questions from falling in the hands of license applicants outside of the test situation. If applicants are permitted to take completed tests with them, the practice in some jurisdictions, all forms of the test will soon be in the hands of the applicant population, allowing the knowledge test to be passed simply by memorizing the answers to each test. If the tests addressed every item of information applicants were expected to know, such a practice would be acceptable. However, this is not the case since available questions

rarely cover every item of information in a manual. The use of automated testing greatly enhances test security.

**Evaluating Knowledge Tests**

Any knowledge test must be evaluated against the purpose it serves. The purpose of a driving knowledge test in licensing is to foster safe operation of automobiles by assuring that drivers possess the knowledge needed to drive safely. It attempts to accomplish this by creating an incentive to secure the requisite information from the driver manual or other sources, and assessing the extent to which the information has been learned. Evaluation of a driving knowledge test includes measures of reliability, validity, and effectiveness. Stated briefly, *reliability* of a knowledge test is the extent to which each administration of the test estimates a person's possession of the body of knowledge that defines the content of the test, *validity* is the extent to which the content of the test truly measures knowledge of safe driving, and *effectiveness* is the extent to which the testing process achieves the objective for which it takes place.

As noted in the introduction, inappropriate application of concepts and techniques borrowed from psychological measurement has hampered the evaluation of driver license tests. Differences between psychometric and license tests based on the reliability, validity and effectiveness are evaluated and will be highlighted in the following discussion.

**Test Reliability**

The reliability of a knowledge test used in driver licensing is the extent to which it accurately estimates the body of knowledge that defines the scope of the test. The questions that make up a license test constitute a small sample of what it is that drivers are expected to know. If the sample does not give a reliable estimate of an applicant's overall knowledge, the test may fail many applicants who really know enough to pass while passing many applicants who do not. The fact that a driving knowledge test has high reliability doesn't mean that the knowledge being measured has anything to do with safe driving; that is validity. However, if the sample of knowledge making up the test does not provide a reliable measure of what a driver knows, it cannot possibly measure knowledge of safe driving.

*Estimating test reliability* - The reliability of knowledge tests, like that of any measurement process, can be estimated by comparing the results obtained from different samples of what is being measured. If the various items making up a test are accurately sampling what a person knows, then scores on individual items and groups of items should give similar results. The larger the sample of questions, the more similar the results should be and the more reliable the estimate of knowledge. If different samples give widely differing results then the score individual applicants receive will depend more upon which questions they were asked than on how much they know.

*Use of alternate forms* – The reliability of the entire test can be assessed where alternate forms of the test are available. If each form provides an accurate measure of total knowledge, then scores on alternate forms should correlate highly with one another.

Reliability can be estimated by having a group of applicants take all forms of the test and comparing test results across the different forms.

*Split-half measures* – Where alternate forms are not available, reliability may be estimated by comparing scores obtained from items making up one half of the test with scores from the other half and statistically projecting the reliability of the entire test. The most commonly used procedure for doing this is the Spearman-Brown Prophecy formula. In estimating "split-half" reliability it's best to compare odd numbered items with even numbered items, rather than the first and second halves of the test, in order to equalize the content of the two halves should the items follow any ordered sequence.

*Item-test relationships* – Kuder-Richardson formula 20 estimates the internal consistency of a test by adding individual item variances (the proportion passing times the proportion failing an item) and subtracting the sum from total test variance to furnish the covariance among items. The KR-20 formula is easily applied with available computer programs and saves both the labor and possible bias involved in splitting a test in half.

*Expressing reliability* – The most common way of expressing reliability is the product-moment correlation between forms or halves of a test. However, limitation of correlation is its sensitivity to test variance; the lower the variance of any two measures being correlated the lower will be the correlation. Once a license test is in use, if the manual is successful in ensuring all applicants possess all or almost all of its content, variance of test scores could become restricted to the point that correlations among test forms are quite low even though the tests might be estimating knowledge very reliably. The interpretation of correlation must take test variance into account. An alternative expression of reliability is the *standard error of measurement*, which provides an estimate of how accurately scores on the test estimates total knowledge. On a test with a standard error of measurement of ±5%, an applicant's score will come with 5% of what the applicant knows about two-thirds of the time. The advantage of using standard error of measurement is that it is relatively unaffected by test variance and its meaning is easily understood.

*Test equivalence* – Various forms of tests may correlate highly with one another, so that people who score the highest on one form also do the same on others, and yet vary in difficulty of the items so as to yield differences in average scores, or in the range of scores. While this type of error can be overcome by converting raw scores to percentiles or standard scores, it is more conveniently avoided by assigning items to forms in a way that will make the different forms approximately equal in difficulty and variance to begin with.

*Test length and reliability* – The most direct way of achieving acceptable reliability is by assuring that the number of items making up each form of the test is large enough to provide reliable samples of what applicants know. A direct way of finding out how large a test must be is by giving the full set of questions available to a group of drivers and

seeing what happens to the correlation and standard error of measurement when the items are divided among more and therefore shorter test forms. Where the correlations begin to drop off and the standard error begins to rise provides a clue to minimum acceptable test length. While there's no fixed number of items for a reliable knowledge test, somewhere in the neighborhood of 40 - 50 items are usually needed for a reliable result.

**Test Validity**

The validity of any test is a function of the extent to which it measures what it claims to measure. Thus, the validity of a driver licensing test is a measure of how well the test indicates that the applicant knows how to be a safe driver. There are a number of means by which the validity of tests is assessed.

*Content validity* - The form of validity most appropriate, *conten*t validity is the extent to which the test truly assesses an applicant's knowledge with respect to a defined body of informational content. Thus, a licensing test for drivers has *content validity* if the test is related to the knowledge or skills necessary for safe operation of the vehicle.

*Discriminant validity* - A test is also considered to be valid if highly experienced drivers generally score higher on the test than do novice drivers. This is called *discriminant validity*.

*Source of content* – In most states, the documented record of what drivers are expected to know is the driver manual. However, the content of the test as a measure of what it takes to drive safely depends upon the source of the content itself. The knowledge requirements specified earlier were derived from a systematic analysis of driving tasks and an evaluation of their criticality to driving safety. A test based upon these requirements would be content valid to the extent that the task analytic process yielded an accurate picture of what drivers must do, and inferences as to the knowledge underlying them were accurate. Until a more systematic process produces a different outcome, the test offers as much content validity as is currently possible.

*Public acceptance or face validity* – To the public taking a knowledge test for licensing, content validity is defined practically by the content of the manual they used in order to prepare for the test. This definition becomes apparent when examiners are called upon to justify questions to applicants who answer them incorrectly. When asked "show me where it says that!" the ability to point to page and paragraph becomes a practical measure of validity. Content validity can be enhanced by seeing to it that items are drawn from the full range of information defining manual content and are expressed in largely the same terms as are used in the manual.

*Predictive validity* - If knowledge plays a part in preventing accidents, then the most knowledgeable drivers should have the fewest accidents. The validity of driving knowledge tests has often been assessed by correlating test scores with indices of subsequent driving to see how well the scores predict who will have accidents, violations or others indices of unsafe driving.

However, predictive validity is less than appropriate for the evaluation of driver knowledge tests in a licensing application for three reasons:

- *Multiple causation* – A serious obstacle to the use of correlation in evaluation measures is that the knowledge measured is among a large number of influences upon driving behavior and therefore upon accidents. Many variables have a much stronger influence upon driving than knowledge and can easily conceal its effect. For example, if drivers with high scores happened to drive more than those with low scores, they could show larger numbers of accidents. Some of the influences such as annual mileage, age, gender or education, can be controlled statistically, at least to some extent. However, many of these confounding influences cannot be measured or even estimated because the effects of knowledge can be difficult if not impossible to isolate.

- *Restriction in variance* – The fact that applicants must pass the license test before being allowed to drive means that the variance of scores achieved by licensed drivers is necessarily limited. If a driving knowledge test were totally effective in leading to the acquisition of knowledge everyone would ultimately score 100% on it, test variance would be zero as would be the correlation of scores with accidents, or anything else. In practice there is little chance that all license applicants will answer all items correctly on any test. However, the restriction in range of scores for those passing the test compounds the problem of multiple causation, further complicating the assessment of tests.

- *Purpose of tests* – As noted in the Introduction, the primary purpose of a driver knowledge test in licensing is not to predict future driving but to improve it by inducing license applicants to acquire the requisite knowledge and to assure that they possess it before they are allowed to drive. It is a quality control measure, functioning in the same manner as a final exam in a school course. The correlation between test score and behavior does not necessarily reflect its ability to change behavior. Evaluating the *effectiveness* of a knowledge test in improving driving safety is an experimental rather than a process of correlation and will be addressed in the next section.

Using prior driving records as a validation criterion avoids the restriction in variance that results from the use of test results for licensing purposes, but not the lack of control over confounding influences. Nor could prior driving records be used to validate tests for novice drivers, who have no such records.

**Effectiveness of Knowledge Testing**

As noted in the previous section, one way in which a knowledge test is expected to contribute to safe driving is not so much by predicting who will be unsafe drivers but by creating an incentive for them to acquire the knowledge needed to drive safely. The test, manual, and other sources of information form a process that can be evaluated for its effectiveness in helping to assure the safety of new drivers. To assess effectiveness in

absolute terms would require having a group of drivers be licensed without having to take the test, and comparing their safety with that of drivers required to take the test. No state has been willing yet to allow novice drivers to operate vehicles without some test of their knowledge in order to conduct such an assessment.

The effectiveness of *improvements* in a manual and test can be assessed by assigning the old and new items at random to large samples of drivers and comparing their subsequent accident and violation records. Several studies have found improvements in license manuals and tests for operators of automobiles and motorcycles associated with lower accident rates. Given the relatively small influence that pure knowledge plays in accident causation, and the many sources of information available to drivers, the effects of improved manuals and tests is expected to be small. The small expense of the licensing process relative to the costs of accidents makes even small improvements in safety highly cost-beneficial. However, the degree of experimental control and the large samples of drivers needed to detect small effects complicates the evaluation of effectiveness.

The complications of distributing different manuals and tests to individual applicants on a random basis, making sure each applicant is tested on the appropriate materials are very great, making control at the level of individual applicants difficult to achieve, although it has been done in several evaluations of license manuals and tests. An alternative is to assign branch offices at random to alternative manuals and tests, following up the subsequent accident and violation records of drivers tested at each location. Control over differences among locations could be achieved by comparing *change* in accident/violation rates for locations in which new manuals and tests were introduced with change (or lack of it) for locations retaining to the original manual.

Experimental evaluation involving control groups involves ethical and legal concerns as well as those of a scientific nature. The fact that the control group is deprived of a process that might reduce its risk of death, injury, or property damage could expose those conducting the evaluation to liability for such consequences. An acceptable defense is that the benefit of the process is unknown prior to the experiment and those in the control group are not being denied anything of proven value. Indeed, the experimental process might prove to be detrimental. The only time that agencies have been considered liable for the negative consequences of control group assignment is when the experiment has been allowed to continue beyond the point that such consequences were known. However, further protection is offered by asking drivers for consent to become part of an experiment, allowing those with objections to opt out. Only those giving informed consent would become part of the experiment and randomly assigned to new versus existing manuals and tests.

**Automated Testing**

Electrical and electronic devices have been developed and

widely used to automate the administration of knowledge tests. While a wide range of automated devices are available, almost all display test questions on a screen and require responses to be registered through some mechanical device. Two basic methods of displaying questions are through *computer-generated images* and through *video images* stored on computer-controlled CD-ROM, DVD, videodisc or videocassette players. Automation offers a number of potential advantages over written testing:

*Processing* — An advantage of automated testing is reducing the labor associated with scoring tests. Since scoring of written tests generally requires only seconds, the labor saved is limited. However, the same automation also permits results of testing to be entered into driver records and stored for statistical purposes. There also exits the possibility of using the same computer-controlled displays to test visual acuity and other visual functions, and even allowing some applicants to enter identifying and personal history information directly into the driver license file.

*Feedback* — Automation allows applicants to be given correct answers after they have responded and their answers have been recorded. While informational feedback contributes to the acquisition of knowledge, the size of the contribution is minimal. Since the items of a knowledge test comprise but a small sample of what drivers are supposed to know, filling in the specific information gaps revealed by any one test administration is but a small step toward overcoming the full range of knowledge deficiencies. Probably the greatest value of feedback is relieving examiners of the need to explain and justify answers to applicants who fail the test and ask to review their results, an activity that can be more demanding of an examiners time than scoring the tests.

*Individualization* — Automation allows different sets of test items for individual applicants. The advantage of such individualization is two-fold. First, it prevents applicants from knowing in advance the specific set of items on which they will be tested, thereby rendering the test largely cheat-proof. Second, it allows the use of "adaptive" sequencing in which the order of items is determined by applicant performance, applicants need not complete the entire test if the responses to the first questions make it probable in passing the entire test close to 100%. The time saved frees up the test equipment for other applicants, thereby making a more efficient use of the equipment. The test could also be terminated as soon as they get enough incorrect answers wrong, although public relations considerations may dictate, allowing them to complete the entire test.

*Imagery* — The electronics of most automated equipment allows detailed static and dynamic images to be displayed in full color relatively inexpensively. This capability is particularly valuable in presenting information concerned with driving. Questions about driving situations can be presented in the same way they occur on the road, allowing complex situations to be addressed without placing demands on verbal skills. If applicants know what to do in actual driving situations they will be able to answer questions correctly, a requirement for valid testing that often cannot be achieved through written knowledge tests.

## SPECIAL APPLICANT CATEGORIES

The requirements that have been described to this point apply to Basic Driver's License applicants in general. Additional requirements are created by the needs of *reading-limited* and *foreign-speaking* applicants.

### Reading-Limited Applicants

License applicants may possess the fundamental abilities needed to operate a motor vehicle safely and yet be unable to pass a driver knowledge test due to (1) inability to read a driver manual and acquire the information needed to pass the test, and/or (2) inability to understand written test questions well enough to pass a driver knowledge test even when they possess the requisite knowledge. This section will address the special needs of applicants whose language limitations are such that their needs cannot be met through the driver manual or driver knowledge test.

### Providing Information to the Reading-Limited

Applicants who are unable to read at a fifth or sixth-grade reading level do not have a source of information available to them, which is needed to meet knowledge requirements like literate applicants are provided. While some reading-limited applicants are able to meet knowledge requirements with the aid of friends or instructors, others either drive unlicensed or manage to pass tests and obtain licenses without really meeting knowledge requirements. Providing reading-limited applicants some way of acquiring information not only eases the burden upon applicants but helps to assure the safety and mobility of the public. Two alternative methods of providing information that have been used effectively with reading-limited applicants are audio and video presentations.

*Audio presentations* — The information needed to fulfill knowledge requirements can be recorded on audio cassettes that can be loaned, rented, or sold to applicants for auditory delivery of information along with printed materials to present pictorial and other content that cannot be adequately communicated through the spoken word. The audio and print material must be designed specifically to meet the needs of reading-limited applicants. Simply reading the text of the driver manual into a cassette will not suffice. Effective learning and retention will require modification of language and sentence structure, integration of the audio and visual presentations, and provision for review and self-testing. The assistance of specialists in teaching the reading-limited should be engaged in preparing cassettes.

*Video presentations* — With the widespread availability of videocassette recorders in homes and libraries, video offers a practical and effective means of communicating to the reading-limited. The increased availability of personal computers opens up the use of DVD, CD-ROM and Internet, media that not only handles visual information but allow interactive forms of instruction to accommodate individual differences in learning and provide for self-testing to help assure learning. However this

medium is not well suited to the use of sound at the present time because the inability to read is often symptomatic of difficulty in handling verbal symbols at all, whether written or spoken, the ability of video to communicate driving-related information in graphic and dynamic form enhances its ability to communicate. Video cassettes and CD-ROM disks may be loaned or rented to applicants, and made available through schools and rental outlets. The availability of a video version of the driver manual has been shown to yield large and significant increases in knowledge levels of reading-limited applicants.

**Testing Reading-Limited Applicants**

License applicants whose reading limitations place them under a handicap when taking a written test should be provided an alternative mode of testing that will yield an accurate estimate of their knowledge despite their limitations. Such alternatives include oral tests, pictorial tests, and audio-visual tests.

*Oral Tests* — Reading-limited applicants may have the written test read to them by an examiner, who also registers their selection of answers. However, there are several drawbacks to oral testing: (1) the spoken word is incapable of handling questions dealing with highly visual content, such as signs and signals, (2) even where reading-limited applicants are able to understand questions, an oral test still places them under a handicap in that they are unable to weight the alternative responses simultaneously in selecting among them, (3) the oral testing process can be influenced by the examiners, who are known to provide aid that is unavailable to an applicant taking a written test, and (4) oral testing is extremely labor-intensive and therefore an order of magnitude more costly than administration of written tests. For these reasons, simple oral testing is not recommended as a means of examining reading-limited applicants. Where it is employed, the pass-fail rates of individual examiners should be reviewed periodically to identify those whose averages deviate sharply from others in the same office.

*Pictorial Tests* — The handicap imposed by the inability of oral examiners to view and review alternative responses simultaneously can be overcome through pictorial representations of each alternative. While each alternative must still be read to the applicant, the pictorials serve as stand-ins for test text. Applicants can consider the alternatives at length and point to the correct answer in order to register their selection. As with written tests, alternative forms of the pictorial tests must be available in order to keep applicants from passing tests by memorizing answers rather than learning the material. While it overcomes one limitation of oral testing, use of the pictorial test is still very labor-intensive and highly subjective. If licensing agencies are willing to abide the cost, the pictorial oral test is acceptable so long as examiners administer it objectively. The pass-fail rates of individual examiners should still be reviewed periodically.

*Audiovisual Tests* — The addition of sound to automated testing provides a means of testing reading-limited applicants. Present day DVD, CD-ROM, videodisc and

videocassette technology allow questions to be presented objectively and understandably in audiovisual form. Moreover, the ability of the visual presentation to display motion helps overcome any limitations of applicants in their ability to infer motion from the spoken word or from still pictures. Because the system is entirely automated, it (1) renders oral tests as objective as written tests, and (2) virtually eliminates the personnel costs of oral testing. An effective audiovisual test possesses the following characteristics:

*Presentation of questions* — The question and alternative answers can be displayed serially on a video display terminal, after which all alternatives can be displayed simultaneously in split-screen form. As the narrative for each alternative response is repeated, the corresponding frame can be highlighted in order to assist applicants in associating answers with corresponding frames. Where the visuals involve motion, the split-screen presentation should select key frames that clearly distinguish one alternative answer from another.

*Registering answers* — Applicants may register their answer either through a special answer pad or by touching a touch-sensitive screen. The latter, although the more expensive approach, is easier for applicants to understand and minimizes errors in registering responses.

*Question review* — Applicants must be given an opportunity to review the question and alternative answers, just as a literate applicant can do on a written test. This can be handled by providing the applicant a means of repeating the question as desired. The number of repetitions should be limited in order to prevent applicants from tying up the equipment for long periods of time while reviewing questions. Where applicants are unable to answer a question in a reasonable length of time, the problem is very likely to lie in ignorance of the correct answer rather than inability to understand the question.

*Summoning help* — Applicants must be provided a means of summoning help if necessary. However, such requests must be discouraged if the advantages of automation are to be fully realized. If applicants allow a long period to elapse, (e.g., 30 seconds) without either answering a question or requesting repetition, a message should tell them how to summon assistance. Examiners must determine whether the request for assistance results from difficulty with equipment or merely ignorance of the correct answer. They should anticipate that applicants who are not adequately prepared will tend to attribute their shortcomings to the equipment.

**Foreign-Speaking Applicants**

The inability to read or speak the English language is not necessarily a barrier to proper motor vehicle operation so long as drivers meet prescribed knowledge requirements and are able to interpret highway signs, signals and markings. It is the responsibility of the licensing agency to assure that these conditions are met before issuing a license. Fulfilling this responsibility imposes special requirements. Given the

importance of mobility to the welfare of the individual, the inability of the foreign-speaking to pass the regular knowledge test will often lead to gaining a license fraudulently or operating a vehicle without one. Therefore, steps taken to accommodate the needs of foreign-speaking applicants will help prevent unqualified drivers from threatening the safety and mobility of the motoring public.

**Presenting Information to the Foreign-Speaking**

A foreign-language version of the driver manual represents the most direct way of presenting information to the foreign-speaking.

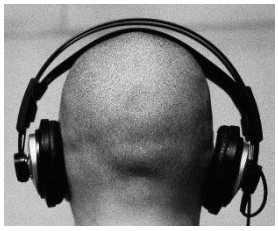
*Preparing materials* — Where the numbers of applicants speaking any one language are large, copies can be printed in volume, the same way as the English-speaking version. However, where the population is small, copies can be reproduced in limited numbers on an as-needed basis. A relatively inexpensive alternative to printed materials with small populations is an audio cassette translation to accompany the regular manual (for graphic material). Cassettes could be made available on a purchase or loan basis.

*Obtaining Assistance* — The cooperation and assistance of nationality- and language-affiliated groups often can be enlisted in preparing materials for their constituency. Such voluntary assistance is beneficial in (1) minimizing costs involved in translating information, (2) resolving issues involving dialect and usage for different countries and regions speaking the same language, and (3) enhancing the credibility of materials among users. Where licensing authorities are unable to assume the burden of preparing a foreign language manual, cooperating organizations may be encouraged to assume the responsibility of printing and distributing materials.

**Testing Foreign-Speaking Applicants**

Foreign language versions of written test forms can be prepared with volunteer group assistance as was suggested for informational materials. Steps that should be taken to ensure the equitable testing of foreign-speaking applicants include the following:

*Alternate forms* — Foreign language versions should be prepared for each of the alternate forms available in testing English-speaking applicants. Where only one form is available, foreign-speaking applicants frequently succeed in memorizing the order of answers.



*Oral testing* — Where foreign language versions of tests are unavailable, or where applicants are not literate in their own language, oral testing may be the only way to assess knowledge. Applicants should ***not*** be permitted to bring their own interpreters, who may provide answers rather than questions. If no examiner speaks the language, arrangements should be made to employ, or obtain volunteer assistance from a reliable independent interpreter.

*Audiovisual testing* — If an audiovisual test is available for examining reading-limited applicants, the addition of foreign language sound tracks will provide a cost-effective alternative to oral testing.

## GUIDELINES FOR SKILL TESTING

The purpose of a skill test, as addressed by these guidelines, is to assess an applicant's possession of the skills that are required to operate an automobile in a manner consistent with the safety and mobility of the motoring public. A skill test cannot predict how safely people actually drive when they are not being tested. The way people behave on the highway is determined to a great extent by attitudes and habits that are seldom revealed in the presence of a licensed examiner. Indeed, research has shown there is no relationship between the driving practices shown on a driving test and those that occur after an applicant leaves the license station. What a skill test *can* do is to force applicants to acquire requisite skills through instruction and practice, and to assure possession of these skills, before they are issued a license to operate unsupervised.

### Skill Requirements

A driving "skill", as defined by these guidelines, is an ability that requires both knowledge and practice for its attainment[1]. The specific driving skills addressed by the guidelines are those that are considered critical for the safety of both the driver and other road users. The guidelines do not deal with skills needed solely for traffic flow, fuel efficiency, or passenger comfort. The skills that are required in safe automobile operation are divided into the following three categories:

*Perceptual* — The ability to perceive characteristics of the many highway traffic environments in a way that permits safe vehicle operation, e.g. judging gaps, identifying hazards.

*Attentional* — The ability to focus and shift attention, e.g. to monitor traffic ahead and to the side in a merge.

*Motor* — The ability to manipulate controls in order to maneuver the vehicle, e.g. ability to rotate the steering wheel in relation to the motion of the vehicle and intended path when turning a corner.

While basic cognitive skills are required to learn and apply facts, procedures, and principles, driving-specific cognitive skills are primarily concerned with navigation rather than safety. Specific skills in each category are summarized on the next page.

Advanced skills, such as those involved in rapid stops or controlling skids, are not among those currently included in driver license testing. Their benefit to safety is

[1] Driving specific skills are distinguished from basic sensory, attentional, perceptual, cognitive and psychomotor abilities. While tests for the latter are appropriate to assessment of individuals who may be deficient with respect to certain abilities, they lie outside the scope of initial license testing. The term "skill" has also been used to mean the *level* of some ability, e.g. to have “a lot of skill.” To avoid confusion, the term “proficiency” will be used to refer to ability level.

open to question, as is the ability of new drivers to master them. However, the introduction of graduated licensing systems, along with the development of devices to permit safe and economical testing of the advanced skills described may change this situation at some time in the future.

**TABLE 3**
**SKILL CATEGORIES**

1. **Attentional Skills**

   *Attention-sharing* — Controlling and maneuvering a vehicle while attending to traffic controls and other road users (search, signaling, space management)

   *Attention shifting* — Shifting attention as needed (ahead, to the side, and to mirrors)

2. **Perceptual Skills**

   *Spatial judgment* — Judging the nature and magnitude of changes in speed and direction of other road users

   *Gap judgment* — Judging the adequacy of gaps when merging, crossing, or entering traffic

   *Distance judgment* — Judging the adequacy of distance of an oncoming vehicle when passing

   *Hazard detection* — Detecting hazards in the characteristics and motion of other road users and in the roadway environment

3. **Routine Motor Skills**

   *Acceleration* — Regulating pedal force to accelerate on level and inclined surfaces.

   *Shifting* — Coordinating clutch, accelerator, and shift lever if manual transmission is used.

   *Maintaining speed* — Regulating accelerator force in order to maintain a steady speed

   *Lane keeping* — Coordinating speed and steering in order to keep the vehicle position within lane on straight and curved paths.

   *Turning* — Coordinating speed and steering when turning corners.

   *Slowing* — Regulating brake and accelerator to reduce speed

   *Stopping* — Coordinating brake, accelerator (and clutch) to bring the vehicle to a stop at a given point

   *Backing* — All of the above in moving the vehicle backward

   *Adjusting to limited traction* — All of the above when operating on slippery surfaces

4. **Advanced Motor Skills**

   *Quick stop* — Regulating braking to stop in the shortest possible distance

   *Swerving* — Turning sharply to avoid a collision while maintaining vehicle control

   *Skid control* — Maintaining directional control of the vehicle during a skid

   *Skid recovery* — Regaining directional control following a skid

   *Evasive Action* — Initiating an appropriate escape action when threatened with a collision

**Testing Requirements**

Skills cannot be measured directly, but are inferred from performance in carrying out various deriving tasks. To yield reliable and valid measures of skills, the performances making up a skill test must meet the following requirements:

*Uniformity* — All applicants must receive essentially the same test, no matter when or where they take it. Applicants with the same ability should have the same probability of passing the test.

*Objectivity* — The scores received by applicants should reflect their performance to the greatest extent possible and be influenced as little as possible by examiners. Two examiners observing the same performance should score the applicant in the same way. Scores should, as much as possible, be reckoned in terms of times, distances, and whether or not a particular response occurred, rather than a subjective judgment, with results totaled such that the applicant's score accurately indicates the level of performance.

*Public Acceptance* — Any skill test must be acceptable to the public in that it must appear to be a valid measure of necessary driving skills, must avoid subjecting applicants to stress or unnecessary embarrassment, and must not expose them to danger.

**Testing Modes**

These guidelines encompass three modes of skill testing:

*Road testing* — Observing and recording the performance of applicants operating their own vehicles over prescribed routes under prevailing traffic conditions.

*Off-street testing* — Observing and scoring the performance of applicants carrying out prescribed maneuvers under controlled conditions in an off-street area.

*Simulation* — Observing the performance of applicants responding to highway traffic conditions simulated through displays and controls that demand of drivers the same skills required in operating an automobile.

The various testing modes are suited for assessment of different skills and require different procedures. Therefore, each is addressed separately.

**ROAD TEST**

The most commonly employed method of assessing driver skill is through examiner observations performed while the applicant operates a vehicle on public streets and highways. The widespread use of road testing as a measure of driving skill reflects its several strong points. First, since it represents a slice of driving itself, the skills assessed in a road test are those required in everyday driving. Second, the applicant's familiarity with the vehicle in which the test takes place allows performance on the road test to provide a more accurate reflection of driving skill than would a test taking place in a simulator. Finally, the test is given in the applicant's vehicle on public streets, avoiding some of the costs incurred in simulator or closed course tests. Testing

costs are limited to those associated with the examiner's time which, of course, is not insubstantial.

Applicant skill is inferred from performance in responding to highway traffic conditions encountered during the road test. The road test itself consists of applicant *performances* that examiners are to observe, the *criteria* that distinguish acceptable from unacceptable performance, *routes* over which the test takes place, *administrative procedures*, and a *method of scoring* performance to determine whether applicants have passed or failed the test

**Purpose of a Road Test**

The purpose of a road test in driver licensing is to assure that drivers have sufficient skill to be allowed to operate a vehicle without supervision. It is not a test of driving practices or habits. Research has shown that there is no relationship between the extent to which drivers demonstrate such practices as signaling, checking the mirror or staying within the speed limit during a road test, and their use of these practices when they are not being tested. The only test performances that correlate with normal driving are those that require the development of skill, such as maintaining the right path in turns and curves, or stopping at the stop line. Checking such non-skilled practices as signaling provides *indirect* measures of skill in that drivers who know they are supposed to signal, having passed the written test, and fail to do so on the road test tend to be those whose vehicle handling skill is so marginal that they are unable to divide their attention between driving and observing various safe driving practices.

In order for the road test to provide a measure of skill, drivers must be informed in advance as to the maneuvers that they will be called upon to make and the specific performances that will be scored. Having to take the test and fail it just to find out what it consists of wastes the time of applicant and examiner. Applicants should not, of course, be advised of specific test routes since it would allow them to rehearse their performance until they can perform the route by rote memory, a luxury they won't enjoy in daily driving.

**Road Test Performances**

Almost all driving performances provide some measure of applicant skill. Performances that require the attention-sharing, perceptual, or motor skills described earlier provide *direct* measures of skill. As noted, other performances that do not require special driving skills for their execution do provide *indirect* measures of proficiency in certain skills by requiring applicants to perform them while applying their driving skills at the same time. Various examples are; visual search, signaling, and obeying traffic signs or signals. Applicants who fail to do these things also tend to be those whose skills are marginal and therefore demand most of their attention.

**Selection of Performances for Observation**

The heart of the road test is the set of performances that examiners are to observe and evaluate. No examiner can observe and evaluate everything an applicant does. The range of performances capable of influencing the public's safety is too great,

and the rate at which they occur during a road test is too rapid to expect examiners to score them all. To be objective and uniform, the road test must identify the performances an examiner is to observe as well as the locations where they are most likely to occur. This is the road testing approach that has been used in several research-based road tests developed for licensing purposes, including the Michigan Driver Performance Measure, the USC Safe Performance Test, the Automobile Driver On-Road Performance Test (ADOPT) and the Commercial Driver License (CDL) Test, as well as the Motorcyclist In-traffic Test (MIT). It is also the approach employed by the California Driver Performance Evaluation (DPE), which forms the basis of the Model Road Test noted in the Preface to these guidelines. Where the examiner's attention is not directed at specific performances, many of them are overlooked. Also since most examiners tend to notice errors more than successful performance, applicants may be scored primarily on what they did wrong, and their chances of failing the test increases with the number of situations they encounter. Tests intended primarily for self-evaluation or improvement is not subject to the same need for objectivity as are tests that determine eligibility for a license.

In deciding what performances should be scored, the following needs to be considered:

*Opportunity* — Situations requiring each performance must occur with sufficient regularity to assure that all applicants are scored on the same set of performances. Looking for performances that depend upon particular traffic conditions or weather conditions tend to be unproductive as well as detracting from the uniformity of the test.

*Objectivity* — Performances that can be assessed objectively are to be preferred over those that require subjective judgment on the part of the examiner. For example, signaling, checking a mirror, or staying in the correct lane can be defined far more objectively than slowing for a "hazard."

*Safety/Mobility* — Performances that cannot be tested without threatening the safety of the applicant, examiner, or other road users, or obstructing traffic, are more appropriately assessed off-street, if at all. For example, assessing the ability of applicants to swerve sharply is likely to threaten the safety of road users, while parallel parking may tend to obstruct traffic.

**Road Test Maneuvers**

Even when the performances to be observed are very limited, no examiner can possibly observe them all. Whether or not examiners observe a particular performance depends greatly upon whether their attention is specifically focused upon it. One way of assuring that attention will be properly focused is by identifying the sequence of performances as they occur in specific maneuvers. Maneuvers that involve a set series of performances include:

- R/L turns without cross traffic
- R/L turns with cross traffic
- straight across traffic
- left turn, oncoming traffic
- negotiating a curve
- lane change
- merge
- exit

- straight driving
- traffic responses

By memorizing the performances associated with each maneuver, examiners will be assured of directing their attention to the right place at the right time. Moreover, structuring an examiner score sheet according to the sequence in which performances occur during a maneuver facilitates recording applicant performance. Research shows that, by concentrating their attention on a specific set of performances at the point where those performances are most likely to be required, will actually allow them to see more than they would if they tried to observe everything.

To achieve the greatest possible uniformity in testing, the maneuvers making up the test, and the number of times each maneuver is to be performed, should be the same for each test administration. Every applicant in every location throughout the state should, to the extent possible, face the same array of maneuvers. Such will not always be possible; there may be no freeway or other location for an angle merge or exit maneuver. Where necessary, other maneuvers will have to be substituted to at least require the same number of maneuvers. For example, a lane change will require signal, search, speed and control performances similar to those required in a merge.

**Road Test Performances**

Within each of the maneuvers, a set of performances can be listed in the order by which they occur during the maneuver. These performances form the basic elements of the road test, and include the following:

- Signal - for turns, lane changes, merges, exits
- Entry position - in turns, curves
- Entry speed - turns, curves
- Full stop - stop signs, traffic
- Stop position - stop signs, traffic lights
- Gap judgment - cross/enter traffic, lane change
- Search - in turns, lane changes, merges, straight driving
- Speed - in turns, curves, merge, exits, straight driving
- Path - in turns, curves, merges, exits
- Lane selection - in turns, straight driving
- Lane position - straight driving
- Following distance - straight driving

**Traffic Dependent Performance**

Traffic conditions along the test route may require performances that cannot be anticipated (e.g., following a vehicle ahead, responding to a pedestrian crossing the street). Because there is no way of knowing in advance just where traffic-dependent situations will arise (1) the attention of examiners cannot be directed toward them to assure objectivity of scoring and, (2) the number and nature of situations will vary from one applicant to another, making it non-uniform. If the responses of applicants to such situations indicate that they are a clear danger to the public, provision can be made for an “immediate failure” (see below). Lesser dangers can be scored, without detracting greatly from test uniformity, by providing one catch-all check for all responses to unplanned traffic conditions.

**Immediate Test Failures**

Examiners should be permitted to fail an applicant on the road test immediately in the event of performance demonstrating a skill deficiency that is sufficiently great that the continuation of the test is not only unnecessary but may place the applicant, examiner, or motoring public in jeopardy. Such performances would include, but not be limited to, running a red light, driving at extremely high or low speed, driving the wrong way on a one-way street or off-ramp, or requiring intervention of the examiner to prevent an accident.

Some jurisdictions fail applicants immediately for violation of any law. However, since almost all road test performances are legally required, the strict adherence to this practice would result in failing applicants for any error, including failure to activate or cancel a turn signal. If the purpose of a road test is to measure skill, then failing on a technicality any applicant that possesses the required skills serves no purpose except to require the state to give, and the applicant to take, an additional unnecessary road test.

**Criteria for Assessing Performance**

Examiners must be provided a means to assess the adequacy of applicant performance. Scoring the road test must include, for each performance, criteria that will allow the examiner to distinguish acceptable from unacceptable performance. In educational settings, where test results are to be used only to guide instruction, a purely subjective appraisal may be acceptable. However, where test results determine issuance of a driver license, the criteria must be objective and uniform. Two applicants with the same skill should receive the same score, no matter who gives the test. When test criteria are subjective, examiners may score the same performances differently, with the result that the unqualified may pass the test and become a danger to the public, while qualified drivers may fail and have to be re-tested unnecessarily.

The establishment of scoring criteria must take into consideration the variety of situations under which performance is observed, across different routes and at different times of day on the same route. The criteria must be broad enough to apply to virtually all conditions under which the test might be given. They must also be relatively simple; an extremely complex scoring system, or one that relies on examiner judgment, will rarely result in a uniform test. Take, for example, signaling a lane change. Simply requiring the signal be given before initiating the maneuver would be objective and simple, but would not assure adequate warning to others. On the other hand, attempting to prescribe precisely when the signal should be given, would require taking account of so many road and traffic conditions as to become extremely complicated, or leaving the question entirely to the judgment of the examiner. While setting a fixed minimum time interval, such as 3 seconds may appear arbitrary, it is uniform and, when communicated to applicants in advance, entirely equitable.

**Route Selection**

All road testing should take place over specified routes. Examiners should not be called upon to make up routes during a road test. Only by designating routes in advance

is it possible to maintain uniformity in testing.

*Number of routes* — Several routes should be devised for each license testing station. Having only a few routes allows applicants to practice driving each route to the point that their performance reflects their memory of the route rather than their general driving skills.

*Route length* — Fifteen minutes of driving in typical urban-suburban settings typically allows for approximately 150 observations of the driver performances making up the maneuvers that have been listed. This number of observations is a minimum for reliable estimation of a driver's skill.

*Route conditions* — Areas characterized by many traffic lights, heavy vehicular travel or pedestrian traffic should be avoided since they can introduce lengthy delays. The number of performances that can be assessed under these conditions is too few for the time spent. Where temporary conditions (e.g., construction/repair) interrupt traffic or change performance requirements, use of the route should be abandoned until normal conditions resume.

*Maximizing observations* — The various maneuvers differ greatly in the opportunities they provide to observe the performances that involve skill. For example, turning, particularly at uncontrolled intersections, permits much more opportunity to assess driver skill than does straight driving. Routes need to be chosen in a way that will result in maneuvers that maximize the opportunities to observe scored performance.

*Setting up routes* — A test route should be viewed as a path between maneuvers. Locations that permit the full array of maneuvers should be selected first and routes planned to interconnect these locations. The art comes in finding enough of the right locations without requiring frequent or long stretches of straight driving, which provides little opportunity to test skill.

*Separating observations* — Locations at which performances are observed need to be sufficiently far apart to allow examiners time to record applicant performance. For example, requiring two turns a block apart may not allow the examiner enough time between the last observation of the first turn and the first observation of the next turn.

*Uniformity of maneuvers* – Although tests routes necessarily differ from one another, a degree of uniformity can be achieved by seeing to it that they all consist of the same number of each type of maneuver, e.g. three left turns across oncoming traffic, three with oncoming traffic controlled, and so on. Gaining uniformity is facilitated by the functional nature of the maneuvers called for. A "merge," for example, can occur on a freeway or, if one is not close enough to the license station, at any location that requires a merge maneuver.

All routes should be driven at different times during the day with applicants in order to check on: (1) total administration time (applicants generally take longer than the person developing the test), (2) previously unnoticed travel restrictions (e.g., "No Left-

Turn 4 - 6 p.m."), (3) examiner overload (too many observations), or underload (stretches where no performance can be observed), and (4) points of possible confusion in giving directions (e.g., a turn at the corner partially hidden by buildings or parked vehicles).

**Administrative Procedures**

The general administrative procedures covering road tests should be thoroughly documented.

*Information to Applicants* — Applicants should be apprised in advance of the performances that will be observed on the test and the criteria that will be used to evaluate them. As noted earlier, applicants should not have to take the test once to find out what skills are being tested. Publicizing the test also enhances its credibility by communicating its uniformity and objectivity. The information can appear in the driver manual or a special handout.

*Communicating with applicants* — Applicants should be dealt with pleasantly and courteously in order to put them at ease and thereby enable them to perform at their best. They should not, however, expect examiners to converse with them since such interaction will interfere with their performance. They should be encouraged to interpret the examiner's silence as an attempt to benefit them, not as a sign of indifference.

*Giving directions* — In guiding applicants during the road test, examiners should adhere to the following:

- Make instructions brief, using non-technical language;
- Employ landmarks that are obvious, and avoid depending solely upon street names, which would not be familiar to applicants who do not live in the vicinity;
- Give the location at which a maneuver is to be made and then the maneuver (e.g., "At the next intersection, turn left"). If the maneuver is given first, applicants may respond immediately;
- Avoid including in the directions any information that is part of the test itself (e.g., applicants should not be instructed as to which lane to enter at an intersection).

**Test Support**

The following items are needed for support of the road testing process:

*Guidance Materials* — Examiners should be provided written descriptions of each route, including the locations at whichspecific maneuvers are to be carried out. Examiners should expect to use route guidance materials only during practice test administrations. It normally takes but a few test administrations to commit the route and the maneuvers to memory.

*Examiner preparation* — Examiners should be trained in administration of the road test and monitored for their ability to provide applicant instructions, observe performance, and interpret scoring criteria. Practice should be provided by having examiners administer the test to instructors, who can make deliberate errors and who can

critique the examiner's administrative procedure and scoring.

*Examiner certification* — Each examiner should administer the test to real applicants over each test route while an instructor or senior examiner monitors the test administration from the back seat. Following the test, the instructor should review and critique the examiner's assessment. This process should continue until the instructor is satisfied that the examiner is able to administer the test correctly.

*Monitoring examiner performance* — The performance of all examiners should be monitored periodically. In addition, the mean scores given to applicants by each examiner should be reviewed. When examiners within a single licensing station report widely divergent mean scores, their testing procedures should be reviewed to discover the source of the discrepancies.

**Test Scoring**

The road test should be administered and scored in a manner such that each individual score reflects applicant skill to the greatest possible extent and the influence of individual examiners is minimized.

**Scoring System**

As with the knowledge test, the road test is scored by aggregating results across all the individual performances. To achieve the *uniformity* that characterizes knowledge tests, road testing must take into account differences among the routes over which the test is conducted. Like different forms of the knowledge test, routes can be made comparable by assuring that they involve the same number of performances and the same pattern of maneuvers. A single passing scorewould prevail no matter where the road test was given.

In some jurisdictions, it may not be possible to render all test routes equal with respect to numbers of observations. Differences in the *number* of performances required can be accommodated simply by expressing the passing score in *percentage* terms. Then, the passing score for any one route can then be calculated by multiplying the total number of observations by the required percentage.

**Recording Scores**

Examiners must be provided a means of recording the performance of applicants as each performance is observed. A suitable score sheet has the following characteristics:

- Every performance to be observed on the road test must appear on the score sheet, both to prompt the examiner and to permit applicant performance to be recorded. To facilitate administration and scoring, performances should be organized on the score sheet by maneuver with performances ordered in the sequence in which they occur within the maneuver, e.g. Lane Change: search, signal, speed, etc.
- Score sheet format should provide space for examiners to record whether the applicant passed or failed a given performance. Since most applicants will pass the majority of performances, the time required to record results can be minimized by noting only those performances an applicant has

failed. Where applicants make no errors on a maneuver, an entry marked "no errors" may be checked by the examiner. Providing this entry helps distinguish truly errorless performance from instances in which the maneuver was not performed or the examiner failed to record an error.

- Space should be provided for entering the total number of correct performances or errors (being fewer in numbers, errors are easier to count). Test forms should also include the passing score. Where the various test routes involve different numbers of performances, the passing score for each route should be either preprinted on the score sheet or provided to the examiner by some other means.

**Evaluating Road Tests**

Road tests, like knowledge tests, can be evaluated for their *reliability*, the extent to which the performances making up the test accurately estimate the applicants full performance ability, *validity*, the extent to which the performances provide measures of the skills that they purport to measure, and *effectiveness*, the extent to which the road testing process improves safety of vehicle operation. Evaluating the reliability, validity and effectiveness of a road test involves the same general considerations and procedures as evaluation of knowledge tests, but with some important differences resulting from differences in the nature of the tests.

**Road Test Reliability**

Just as the reliability of a knowledge test is how well it estimates an applicant's knowledge, the reliability of a road test is a function of the extent to which the scores given to an applicant on any one administration of the test estimate the overall performance capability of applicants. Various administration of a road test involves using different routes, which are somewhat analogous to the different forms of a knowledge test. Each route samples the skills required in driving, just as each knowledge test form secures a sample of knowledge. However, road testing introduces concern for another aspect of reliability, which is the reliability of examiners in observing, interpreting, and recording performance. Unlike marks on a multiple choice answer sheet, the scores assigned by examiners represent subjective judgments, influenced by how much and what examiners observe and how they interpret it. The total measurement reliability of the road test is therefore a joint function of the *examiner reliability* and the *route reliability*. While it is the total reliability that counts, assessing the examiner and route reliability separately provides insight into the sources of unreliability and guides efforts to improve the test.

The way reliability of road tests is generally assessed is by comparing results gained from having the test administered to the same applicant by different examiners and over different routes. If the test is measuring true performance capability, the results should be highly similar. If they are not, then the test may be measuring characteristics of the examiners and routes as much or more than it is measuring characteristics of the applicant. Similar results do not guarantee that the performance capability of the test

relates to safety; that is a matter of validity. Reliability is a necessary but not sufficient condition for validity.

**Examiner Reliability**

The examiner reliability of road tests is estimated by having two or more examiners score applicants over the same routes (two examiners is generally the most that can see the driver well enough to score performance) and assessing the extent to which the different examiners agree on the relative scores assigned to the same applicant. It can be expressed the same way for a knowledge test, by calculating the standard error of measurement, and making sure that it is no more than a few percentage points. In addition to assessing reliability of the entire road test, the reliability of individual performances making up the road test, such as visual search, signaling, or path, needs to also be examined to identify those aspects of performance with which examiners may be having difficulty. While the reliability of individual performances will be lower than those for entire tests, comparing across multiple examiners scores will help point to those whose scoring criteria might be improved to raise examiner agreement.

A road test can show high agreement or correlation among examiners as to who are the best performers but show substantial differences in average scores that different examiners give to all applicants. Some examiners are just more sensitive to applicant's errors than others and give generally lower scores than others, overall or with respect to specific performances. Equivalence can be assessed by comparing mean scores or pass-fail rates assigned to the same applicants by the different examiners. Unfortunately, examiners cannot be readily "equated" for difficulty by generating percentile or standard scores for different examiners in the same way as alternate test forms.

Low examiner agreement and equivalence, for the overall test or at the level of individual test performances, can be addressed by seeking greater objectivity in scoring criteria and providing the instruction needed to assure that examiners can and do adhere to the scoring criteria. It is possible that certain road test performances simply cannot be dealt with in a way that yields acceptable examiner performance. In such cases it may be necessary to eliminate the performance as an element of the road test. If examiners cannot agree upon the scoring of a performance, and there is no way of knowing who is correct, test results will not be valid and there is no point to include an invalid measure in a test, regardless of how important the behavior may be, for all applicants.

The fact that a road test itself is shown to be generally reliable does not mean that it will be reliably administered by all examiners in operational use. For this reason, individual license stations should monitor the overall pass rates or mean test scores for their examiners in order to identify any that are significantly above or below the average for that station. In addition to overall test results, the pass rates and average scores for each of the performances need to be examined to detect examiners who have certain "pet" performances where they are more likely to see mistakes than others.

Except for examiners who deal with special cases (e.g. elderly drivers, certain high schools) different examiners should handlethe same general kinds of applicants and have similar distributions of scores. Examiners whose test results are deviant are candidates for close monitoring and possible retraining.

**Route Reliability**

Various road test routes, like different forms of a written test, can yield varying results. Tests given in city traffic can be expected to expose drivers to greater chance of error than tests in suburban areas when there is little traffic. Route reliability can be assessed at the same time as examiner reliability by having the same applicants scored by the same examiners over pairs of routes and comparing the scores given by each examiner over the different routes. The correlation across routes should be relatively high, and the standard error of measurement for estimating performance on one route from the other should be within a few percentage points. Performance on different routes may correlate with one another and yet lackequivalence, owing to differences in the conditions that lead to error, similar to differences in the difficulty of knowledge test forms or differences between "hard" and "easy" examiners. Therefore in addition to standard error measurement, the mean scores compiled on the different routes should be compared. It is quite possible that substantial differences will arise within the same route during different times of the day. In some locations, the density of rush-hour traffic may lead to high error rates. If so, routes might be revised as to location or hours of the day.

As with examiner reliability, the route reliability should also be examined at the level of individual performances to see which aspects of the test may be contributing to the error that arises. Where variation in routes leads to large errors of measurement or large mean differences, the route selection criteria needs to be revised in order to yield routes that are more comparable. It may prove impossible to overcome route-to-route differences for certain performances, in which case it may be necessary to eliminate the performance measure from the test in the interests of validity and equity. The fact that a road test shows acceptable route reliability in general doesn't mean that all individual routes will give comparable results. Equivalence among examiners can be assessed in the same way as examiner equivalence, by comparing mean scores. Within each license station, mean scores compiled on the various routes should be compared with one another from time to time. If all examiners are using the same routes, differences in means scores would be expected to reflect differences in the opportunities for error. Checking for route equivalence is particularly important after changes in road characteristics, or events that might alter traffic patterns.

**Total Reliability**

The total reliability of the road test is a function of both examiner and route reliability. Where pairs of examiners give tests over two routes, it is estimated by comparing the scores given by one examiner on one route with those given by the other examiner on the other route. It corresponds most closely to the reliability of a road test in actual use, where different applicants are tested by different examiners over different

routes. Again, reliability can be expressed in terms of the measurement error in estimating scores on one route with one examiner from scores on a different route with a different examiner. The inability to standardize examiners and routes to the same degree as written tests results in a reliability considerably lower than those of written tests.

**Road Test Validity**

The considerations and procedures involved in assessing the validity of road tests parallel those of knowledge tests. The methods by which performance tests are administered and scored must provide valid measures of the ability to operate an automobile in a manner that protects the safety and mobility of all road users. Applicants who lack the skills to meet these conditions should not be capable of passing the test. Like knowledge tests, the validity of road tests is not easily assessed by measuring their ability to predict who will have accidents; applicants predicted to have accidents don't get todrive unsupervised. A road test's content validity can be assessed by the extent to which the driver performances have been identified as related to safety of operation. The performances that were identified earlier derive from a systematic analysis of the performances that make up safe driving and the abilities that prior analysis of driving tasks, and research into driving skills, appear to identify as important to driving safety.

A measure of road test validity can be gained by comparing the scores of experienced drivers with those of novices. If one assumes that skills develop with practice, then drivers who have had a lot of experience should outscore rank beginners. In order for experience to manifest itself in skill, the experienced drivers must know the performance on which they are being scored and motivated to perform to the best of their ability. Highly experienced and skillful drivers can make careless errors, which may reflect adversely upon their everyday driving, but not their skills.

**Effectiveness of Road Tests**

The purpose of road tests is to prevent accidents by assuring that drivers meet minimum skill requirements before being licensed to drive unaccompanied. Their absolute effectiveness in this regard could be evaluated experimentally by comparing accidents of driver's licensed with and without road tests, or some other measure of skill. In practice the effectiveness of road tests, like that of knowledge tests is unlikely to be assessed in this way due to the unwillingness of jurisdictions to issue licenses without some demonstration of at least the basic ability to drive a car. However, improvements in road testing could be evaluated through controlled experiments in which drivers licensed under competing road tests are compared for accidents and violations subsequent to licensure. Thus far no such controlled evaluation road test effectiveness appears to have been conducted. Acceptance of improved road testing procedure has stemmed primarily from the accepted validity of their content and, in some cases, their demonstrated reliability.

**OFF-STREET TESTING**

Off-street skill testing has been used primarily for three purposes: (1) initial screening for minimum skill levels before applicants are exposed to the potential

hazards of road testing, (2) allowing for certain vehicle control skills to be assessed more efficiently than is possible in an uncontrolled road environment, and (3) permitting assessment of emergency skills not safely assessed in a road test. Since each of these functions involves a somewhat different array of test requirements, each will be addressed separately.

**Pre-test Screening**

The ability of applicants to control the vehicle may be so marginal as to make road testing a hazard to applicants, examiners, and the motoring public. While the number of dangerously unqualified applicants may be extremely small, the interests of safety are best served by identifying such applicants before a road test commences. This can be handled by contriving the road test to start in a parking area such that several turns and stops are required before entering the road. If a suitable parking area is not available, the test can begin on a lightly traveled side street.

**Assessing Vehicle Control Skills**

Off-street tests have been used instead of or in addition to road tests. The ability of applicants to handle a vehicle can be assessed more accurately in the off-road environment than on the road owing to the ability to: (1) require maneuvers that are more demanding and therefore more revealing of skill levels, (2) measure responses more precisely (e.g., stopping distance), and (3) have completely standardized test characteristics, thereby permitting collection of more uniform and reliable data. Where the necessary facilities are available, off-road testing is also economical, requiring less examiner time to obtain reliable results. Off-street skill testing is widely used in licensing automobile, motorcycle and commercial vehicle operators.

Basic vehicle driving skills that are testable off-road are those involved in accelerating, (including shifting gears), braking, turning corners, and backing. Exercises capable of assessing these skills have been devised using stanchions, traffic cones, painted lines, and in some facilities, traffic control devices such as stop signs, yield signs, and traffic lights. The shortcoming of the off-street test is its inability to measure skills involved in such traffic-related performances as merging, changing lanes, following, and judging gaps, as well as lane keeping, handling curves and approaching turns at highway speeds. As yet there is no research evaluating the relative merits of on- and off-street testing in assuring ability to drive safely.

**Emergency Skills Testing**

The off-street environment offers an opportunity to assess skills in carrying out emergency maneuvers without interference from or risk to other road users. Important emergency maneuvers include maximum braking, evasive steering, and skid recovery. None of these skills are currently part of automobile license testing. Barriers to implementation include cost of needed off-street testing equipment and facilities, potential danger to applicant, and questionable relevance to initial licensing. As yet, no research has established the

relationship between such testing and safety of operation. Instruction and testing in emergency braking and swerving of motorcycles has demonstrated an accident reduction potential. While instruction in handling automobile skids has been associated with elevated accident risk in certain regions, there is no way of knowing whether it makes drivers less safe, or just encourages more driving under conditions in which such skills are required. Certainly, the value and feasibility of testing emergency skills has not been sufficiently well established to make such testing a part of licensing for automobile drivers.

**Evaluating Off-Street Testing**

Evaluating off-street testing as a mode of skill testing (not just the start of a road test) imposes some special requirements.

**Reliability**

The issue of reliability is generally less crucial to off-street than on-street testing for two reasons. First the high degree of objectivity in scoring minimizes the effect of differences among examiners. Second, the fact that there is just one set of performances required removes variance associated with the varying routes and traffic conditions that reduce the reliability of road tests. The variations encountered in repeated administrations of the same test are almost entirely those resulting from inconsistency of the drivers themselves. Procedures for assessing examiner and sampling reliability of off-street tests parallel those of road testing, i.e. having at least two examiners score the same drivers over at least two administrations of the test (any more than two can introduce a practice effect that masks measurement of driving skill with skill in handling the test course). Examiner and sampling reliability are measured in largely the same way as with road tests, by comparing the scores of two examiners during the same administration and by comparing the same examiners scores across two administrations of the test. Total reliability would be the relationship between the scores obtained by one examiner on the first administration and the scores of the other examiner on the second administration.

**Validity**

The content validity of off-street testing is determined by the degree to which the maneuvers performed call for the same skills as are required on the highway. Obvious limitations are the inability to test for skills required in (1) moderate to high speed maneuvers such as merges and lane changes, (2) interacting with traffic, such as judging gaps and following distance, and (3) handling the vehicle while performing other activities, such as checking mirrors, signaling, and watching other road users. The validity of off-street tests will depend upon the extent to which the performance of tests of off- street provides accurate estimates of the full range of on-street performance. This aspect of validity can be assessed by administering both off-street and on-street tests to a representative sample of applicants and comparing results, which use the road test as a criterion in

evaluating the off-street test.

**Effectiveness**

As long as *some* measure of driving skill is to be part of the licensing process, the question of effectiveness is not whether use of an off-street skill test leads to a reduction in accidents, but whether a particular off-street test is more effective in doing so than another off-street test, or a road test. The fact that an off-street test might not be a highly valid measure of total driving skill, as evidenced by its correlation with a road test, doesn't necessarily mean that it is less effective in reducing accidents caused by inept driving. If driver preparation were to focus on the skills most critical to safe operation, and measure those skills precisely, it might lead to safer driving than the more encompassing road test. Conversely, just because an off-street test correlated highly with a road test among a group of drivers doesn't mean it will be effective in fostering the same array of skills as the road test among new drivers. In short, use of an off-street test as a substitute for a road test must be assessed through an experimental comparison and not simply through the correlation between the two types of measures.

**SIMULATION**

The limitations of a road test as a measure of skill has stimulated interest in simulation. The potential benefits of simulation over road testing in the assessment of driving skills include:

S*cope* — in a few minutes, an applicant can be confronted with an array of highway traffic situations that it might take days or weeks to encounter on the road,

*Uniformity* — every applicant can be presented with the same situation, or situations that have been equated for difficulty,

*Automation* — the examiner performance recording, scoring and debriefing functions can be carried out automatically, and

*Safety* — applicants can be presented with hazardous conditions to which examiners may be reluctant to expose to an unlicensed driver.

Thus far the use of simulation has been limited to experimental applications. Despite its benefits, simulation is unlikely to serve as a substitute for a road test in licensing. First, neither the public or licensing agencies are likely to accept the idea that a license should be issued or denied without some demonstration of an applicant's ability to drive a car. Second, simulation devices involve cost whereas the road test is conducted in the applicant's vehicle. Yet, as an adjunct to the licensing process, low-cost forms of simulation may have potential benefit in pre-screening drivers to (1) avoid testing unprepared and potentially dangerous applicants, (2) identify renewal or out of state applicants who may require road testing, (3) to guide examiners in deciding on the nature and length of road testing, and (4) to help pinpoint the source of deficiencies among driver's performing poorly on the road test.

Simulators appropriate to assessment of driving skills fall into two categories; *interactive*, in which drivers respond to simulated highway traffic scenes while the scenes change as a function of what the driver does and *non-interactive*, in which the simulated scenes are prerecorded and

remain the same no matter what the driver does. Each form of simulation is suitable for testing different skills.

**Interactive Simulation**

An interactive simulation is one that people can actually "drive" in that simulated driving scenes change realistically as the driver operates the simulated controls. Simulated motion can be created in two ways:

(1) driving a camera or optical pick-up along a 3-dimensional model of a highway environment and displaying the changing image in front of the driver, and

(2) generating images by means of a computer and displaying them on a terminal. With the advances of computer technology, the latter has become the predominant type of system.

The present-day interactive simulators reproduce vehicle dynamics with very high fidelity, allowing their use to teach and test for vehicle control skills — the routine skills required in accelerating, lane keeping, and braking as well as the emergency handling skills required in swerves, quick stops, and skid recovery. However, simulation truly capable of duplicating the complexity of the highway traffic environment, and changing realistically as a function of the driver's responses, is extremely expensive. Given the economy that prevails at the time these guideline are prepared, the prospects of introducing interactive simulation into the basic license testing process are not encouraging. However, with the rate at which technology is advancing, these prospects could change, and developments bear watching.

**Non-Interactive Simulation**

In non-interactive simulation, drivers respond to recorded images of the scene ahead of the vehicle and that afforded by its mirrors. Since the images are prerecorded, the scenes do not respond to what the driver does; thus this type of simulation is unsuitable for teaching or testing vehicle control skills. Their use in license testing would be limited to testing knowledge and perceptual skill.

*Knowledge* – As a knowledge testing device, simulation has the advantage over written tests of being able to get an answer without having to ask a question. While a written test can determine whether an applicant possesses certain information, it cannot assess ability to recall it when necessary; just asking the question provides a cue to recall, as does the correct answer appearing as one of the alternative responses. Simulation, like driving, can present situations requiring application of information without alerting the applicant to the need to recall information or providing cues to the information itself.

*Perceptual skills* – The media that lend themselves to non-interactive simulation for licensing applications, CD-ROM, DVD, videodisc and videotape (use of film is now largely confined to group educational settings) allow a dynamic presentation of driving scenes with sufficient fidelity to assess the perceptual skills involved in judging gaps, closure rates, and in the detection of hazards. At present, these non-interactive media provide the only means of duplicating the complexity of the highway traffic environment within

acceptable cost. The purpose of vehicle controls in non-interactive simulation is only to provide a means by which the drivers can register what they have perceived, such as coming off the accelerator or braking when they perceive a hazard or pulling back into a lane where passing distance is inadequate.

The driver can be told to expect that scenes will not change as a function of their responses. And while, the driver's response cannot alter the driving scene, it can be designed to cause the scene to end and thus prevent conflict between what the driver does and the way the scene responds.

Because operation of the vehicle control serves simply to register a correct answer, the use of actual vehicle controls is not really necessary. License applicants can register responses by means of a joy stick, answer key, or touching the screen, thus allowing the benefits of simulation to be gained largely from the same equipment employed in automated knowledge testing. The complexity of traffic scenes that can be portrayed through video is somewhat limited by the low resolution of images capable of being presented by off-shelf video equipment. Advances in "high definition" video promise to yield substantial improvements in resolution.

**Evaluating Simulation**

The reliability of any simulation measure would be assessed in the same way as knowledge tests, through the correlation among different samples of the performances being assessed. The validity of simulation in basic vehicle operating skills could be measured through correlation with performance of the same basic maneuvers in a real car. The knowledge and skills that involve response to stable aspects of the driving environment – road characteristics and traffic controls –can be validated in the same manner. In the validation process, performance in the car should be recorded with the aid of instrumentation such as video; human observation is likely to furnish a validation criterion that is less reliable than the simulation being validated. However the knowledge and perceptual skills involving other road users, such as knowing appropriate following distance or recognizing pedestrian hazards, would require conditions that can't be stabilized in the real world. Here, content validity must be established through analysis relating the simulated conditions and required behaviors to those found in driving.

The effectiveness of simulation as a part of the licensing process would have to be evaluated against the specific functions it is intended to serve. Where it is intended to add to improve the ability of the licensing process to reduce accidents, its effectiveness could be compared with that of licensing without simulation in a random experiment.

## GUIDELINES FOR MOTORYCLE KNOWLEDGE AND SKILL TESTING

This section provides some considerations that may be utilized for the development of motorcycle operator knowledge and skills tests.

The development of a motorcycle operator's manual, a motorcycle specific knowledge test and a skills test for motorcycle operators requires that a list of knowledge and skills elements first be developed. However, these lists have not been validated.

The following is a draft list of knowledge and skills elements; however, this list has not met the requirement for community consensus. This list is provided solely to assist motor vehicle administrations in developing their own list of elements.

### Knowledge Elements and Domains for Motorcycle Knowledge Testing

This list provides a sample of knowledge test elements and domains for a motorcycle specific knowledge test and operator's manual. It is assumed that applicants taking a motorcycle specific knowledge test have already passed the basic knowledge test for automobile drivers, which includes signs and rules-of-the-road.

***Vehicle Inspection and Care***
- Pre-ride Inspection

***Protective Gear***
- Conspicuity
- Comfort
- Protection

***Risk Assessment***
- Rider Responsibility

***Motorcycle Controls***
- Location
- Operation

***Vehicle Control***
- Getting Underway/Clutch Control
- Balance and Direction Control
- Shifting
- Braking

***Lane Positioning***
- Visibility (see and be seen)
- Space Cushioning
- Avoiding Surface Hazards
- Escape Routes

***Negotiating Curves***
- Speed Control
- Line Selection
- Visual Directional Control
- Counter-steering for Cornering

***Emergency Maneuvers***
- Stopping Quickly
- Obstacle Avoidance
- Vehicle Failures

***Special Riding Situations***
- Carrying Passengers and Loads
- Roadway Characteristics
- Weather Factors (rain, snow, wind)

***Impairments***
- Alcohol and Other Impairing Drugs
- Distractions
- Fatigue
- Temperature Extremes

**Skill Elements and Domains for Motorcycle Skills Testing**

This list provides a sample of skills test elements and domains for motorcycle skills or on-road testing and the operator's manual.

***Preparation***
- Vehicle Inspection
- Motorcycle Controls
  - identification
  - operation
- Motorcycle Riding Gear

***Vehicle Control***
- Getting Under Way and Riding Slowly
  - smooth clutch control
  - balance and coordination
    - using foot/feet balance
    - visual directional control

***Vehicle Operation***
- Shifts Smoothly
  - no missed shifts or gear grinding
- Maintains Directional Control
- Uses Both Brakes
- Use of Mirrors and Head Checks
- Gap Selection
- Prevailing Speed

***Lane Positioning***
- Visibility
  - being seen
  - Seeing others
- Lane Protection
- Space Cushioning
- Escape Route
- Surface Hazards
- Stop Position in Lane

***Turning***
- Visual Directional Control
- Speed Management
  - lane/boundary violations

***Emergency Situations***
- Quick Stop
- Obstacle Avoidance

For the development and evaluation of operator manuals, knowledge tests and skills/road tests refer to the earlier sections of this document.


AAMVA