INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University Microfilms International

300 N. Zeeb Road
Ann Arbor, MI 48106

8424472

Rudisill, Michael Davis

DEVELOPMENT AND EVALUATION OF AN OBSERVATIONAL MEASURE TO EVALUATE IN-CAR PERFORMANCE OF MICHIGAN DRIVER EDUCATION STUDENTS

Michigan State University PH.D. 1984

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

Copyright 1984

by

Rudisill, Michael Davis

All Rights Reserved

DEVELOPMENT AND EVALUATION OF AN OBSERVATIONAL MEASURE TO EVALUATE IN-CAR PERFORMANCE OF MICHIGAN DRIVER EDUCATION STUDENTS

By

Michael Davis Rudisill

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Administration and Curriculum

1984

© COPYRIGHT BY

MICHAEL DAVIS RUDISILL

1984

ABSTRACT

DEVELOPMENT AND EVALUATION OF AN OBSERVATIONAL MEASURE TO EVALUATE IN-CAR PERFORMANCE OF MICHIGAN DRIVER EDUCATION STUDENTS

By

Michael Davis Rudisill

This study dealt with the development of an observational measure for evaluation of in-car performance of Michigan driver education students and the determination of the measure's reliability characteristics. This instrument was designed for use by the Michigan Department of Education to determine the effectiveness of driver education programs in the state.

An integral part of the study process was the design of a test route that would yield the situations to observe and record the driving performances stipulated by the Michigan Department of Education's in-car performance objectives. Also involved were the design of an instrument that was concise and definitive enough for the raters to use efficiently, design and implementation of a training program for raters, development of a counterbalanced design for rater and subject assignment, and the statistical treatment of the data. Analysis of variance and Pearsons' Product Moment correlations were used to determine statistically the reliability

characteristics of the test.

The study addressed the following research hypotheses:

1. That there would be differences in difficulty among the items in the test. The F ratio was significant beyond alpha .01, indicating that there was a difference in item difficulty.
2. That run administrations would not affect driver performance scores. The findings were not significant, suggesting that performance was stable over time.
3. That subjects' driving performance scores would not vary according to items interacting with the time of test administration. The finding was not significant, suggesting there was no significant interaction between test items and run administration.
4. That a positive relationship would exist between true driver performance scores and observed driver performance scores. Three methods of analysis were used. The correlation coefficients were .957, .937 and .730.
5. That a positive relationship would exist between raters on measures of sum, search, speed control, direction control, familiarization, and signs. The interrater reliability for pairs

of raters ranged from .49 to .83 on test components. The overall test had a reliability coefficient of .86 for pair one and .83 for pair two.

DEDICATION

I would like to dedicate this study to my parents, Carl and Grace Rudisill, for early guidance and support; to my wife Jane, for her sacrifices and support; and to my daughter Michelle, who frequently said, "Daddy, do your homework."

ACKNOWLEDGEMENTS

The completion of this study would not have been possible without the support and encouragement of a number of people.

I would like to thank the members of my doctoral guidance committee: Dr. Robert E. Gustafson, Chairman, Dr. Ben A. Bohnhorst, Dr. William D. Frey, and Dr. Robert O. Nolan. A special thanks to Gus and Bill for their extra efforts.

I would also like to acknowledge Mr. Frederick E. Vanosdall, Dr. MaryEllen McSweeney, and Dr. Kara Schmitt for their valuable support during the planning stages of the study.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION

Michigan has long been recognized as a leader in driver education. In 1955, Michigan was the first state to pass legislation requiring local school districts to provide driver education. Although recognized as a leader, Michigan had not conducted any systematic research concerning the effectiveness of driver education programs. In 1975, the Michigan Department of Education submitted a proposal to the Michigan Office of Highway Safety Planning to evaluate the effectiveness of driver education programs in the state. The Department of Education was successful in receiving a three-year grant from the Michigan Office of Highway Safety Planning. This request for funds was initiated by four primary factors.

First, in 1972, the Michigan State Legislature requested the Superintendent of Public Instruction to report on the effectiveness of driver education programs in the state of Michigan. Four questions were presented in an effort to make this determination:

1. Do current driver education programs insure that the student acquires the knowledge and skills necessary to pass successfully the state driver licensing examination?

2. Is one type of driver education program more effective than another in providing students with the knowledge and skills necessary to successfully pass the state driver licensing examination?
3. Does successful completion of a driver education program have a positive impact on road safety?
4. Is there any evidence to suggest that one type of driver education program is more effective than another in terms of positive impact on road safety?

The 1972 study was incomplete. The study, using questionnaires and interviews, concentrated on question four. Question three was difficult to answer due to the lack of an adequate control group. The 1975 study was, in part, an attempt at answering the general concepts of questions one and two.

Second, in 1974, the state Legislature was requested to increase the amount of driver education reimbursement. The Legislature refused to increase the $30.00 per student reimbursement until proof could be presented that driver education programs were effective.

A third factor prompting the study was the recognition of a trend for some states to favor commercial driver education over public school driver education programs. No studies had been conducted in the state of Michigan to indicate that one type of program was better than the other.

The fourth factor leading to the initiation of the 1975 study was the recognition, by Michigan Department of Education staff, that there were no objective means available to schools and teachers to evaluate the effectiveness of their programs. In response to this recognition of the need for objective criteria to evaluate the effectiveness of driver education programs, the Department decided to develop minimal performance objectives to be used as the basis of instruction. The Department then decided that the most practical approach to evaluating the effectiveness of Michigan driver education programs would be to measure the students' attainment of these objectives.

The first year of the project was spent on the development of driver education performance objectives. The performance objectives were reviewed by 225 driver education teachers, specialists and experts. Revisions were made on the basis of these reviews. The performance objectives were then distributed to driver education teachers throughout the state of Michigan as guidelines to enhance uniformity of course content.

The second year was spent developing a written test, based on the performance objectives, to measure the driver education students' classroom performance. The test was administered to approximately 200,000 students throughout the state. The students tested were from private, parochial, commercial and public schools.

The third year was spent on the development and evaluation of an instrument, based on the in-car performance objectives, to measure the in-car performance of a random sample of Michigan driver education students.

The Problem

Statement of the Problem

No instrument existed that was designed to measure the Department of Education's in-car performance objectives. Consequently, a new instrument would have to be developed for the in-car evaluation phase of the project. The Michigan Department of Education, recognizing the extensive work done by the Highway Traffic Safety Center at Michigan State University in the area of driver performance measurement, asked the Center for assistance in developing a measure of in-car performance of driver education students. The writer, having interest and experience in driver performance measurement, agreed to assist the Department of Education by developing an observational measure to meet the needs of the Department and to determine the reliability of that measure.

In order to make it practical, this phase of the project had to be divided into more manageable parts. One integral part of this process was the design of a test route that would yield the necessary situations to observe and record student driving performance as stipulated by the in-car performance objectives. A second major component of the process was the design of an instrument that was concise, thorough and definitive enough to be easily manageable by the raters observing and recording driver performance. The third integral part of the study was the design and implementation of a training program for the raters who would be observing and recording the driving performances.

This study dealt with the development of the observational measure for evaluation of in-car performance of Michigan driver education students and the determination of the measure's reliability. This instrument was to be used by the Michigan Department of Education to determine the effectiveness of driver education programs in the state. The results of the effectiveness study will be presented to the state Legislature in order to comply with the Legislature's request for proof of program effectiveness before granting an increase in the reimbursement to school districts for driver education expenditures.

Purpose of the Study

The purpose of this study was to develop an observational measure to evaluate in-car performance and to determine the reliability of the observational measure. It was important to ascertain which of the observational measures provided by the instrument were reliable, which were not, and under what conditions. Reliability and validity had to be determined if the evaluation of driver education program effectiveness was to be deemed significant and beneficial.

To be responsive to the concerns for reliability, it was necessary to determine internal consistency or the reliability of items measuring the same objective. Stability in time, with respect to the times of administering the instrument to the same subjects, was another concern. A major concern in regards to reliability was stability with respect to raters. Therefore, it was important to determine that there was positive agreement between raters.

The initial requirement for testing new instrument procedures is the determination of the various reliability characteristics. In other words, will the scores the test yields, across raters, across administrations and across the items, be consistent. In keeping with this concern, this study attempted to test the following hypotheses.

Hypotheses to be Tested

1. It was hypothesized that there would be differences in difficulty among the items in the driver test.
2. It was hypothesized that run administrations would not affect driver performance scores.
3. It was hypothesized that the subjects' driving performance scores would not vary according to items interacting with the time of test administration (i.e. between run 1 and run 2).
4. It was hypothesized that a positive relationship would exist between true driver performance scores and observed driver performance

scores (internal consistency).

5. It was hypothesized that a positive relationship would exist between raters on measures of sum, search, speed control, direction control, familiarization and signs (interrater).

Significance

Accountability is a major concern in all facets of driver education. Administrators, driver educators, parents and public officials are concerned with the accountability of driver education. In addition to being concerned with program offerings and outcomes, they are also concerned with cost effectiveness. This study may very well provide the framework and information necessary to answer the concerns for accountability. More specifically, it may provide the necessary information to answer the concerns of course offerings, program effectiveness and the relative efficiency or cost effectiveness of various types of driver education programs. It may also have implications as to desirable instructional materials, teaching techniques and teacher preparation. It can also serve as a means to reevaluate the existing performance objectives. The study provides a means of evaluating and possibly improving driver education nationwide, as well as in the state of Michigan.

This study more immediately provides the Michigan Department of Education with the means necessary to gather data to present to the Michigan Legislature in an effort to procure an increase in the reimbursement allotment to local school districts for the funding of driver education programs.

Limitations of the Study

This study was limited to an examination of the reliability of an observational measure and its use by trained raters. The observational measure was designed to measure only driving performance specified by the performance objectives developed by the Michigan Department of Education. The observational measure did not include those objectives calling for atypical or hazardous situations.

The selection of raters was limited to volunteers who could arrange their schedules to be available during the time frame of the study. The training program for the raters was designed to meet the time frame in which the raters were available.

This study made no attempt to compare different programs or instructors. No attempt was made to compare teaching experience to rater performance.

Methods of Procedure

Tasks

In conducting the methods and procedures of this study, various tasks were identified and completed. A route was designed to be representative of typical driving environments that yielded traffic situations requiring the driver to display performances stipulated by the performance objectives. The route was divided into areas of observation and areas of recording.

Dividing the route into observational and recording areas conttibuted to the design of an instrument that was comprehensive yet manageable by trained raters.

A training program was prepared for the raters consisting of lecture and field exercises that involved actual observation and recording of practice subjects' driving performance. Vehicles, equipment and classroom facilities were procured for use in the training and data collection phases.

Subjects, having recently successfully completed their driver education course, had to be identified and randomly selected for the training and data collection phases. The subjects used during the training program for raters were not used during the data collection phase. Data were collected on the observations of the subjects used during the training session to determine rater agreement.

The data collection phase immediately followed the 26-hour training program for the raters. Data were collected on 30 subjects. The actual data collection involved seven and one-half days. Statistical treatment was applied to the data for analysis.

Basic Assumptions

In conducting this study, some basic assumptions were made. The assumption that the training was sufficient to insure standardization of procedures to be used by the raters was made on the basis of observing rater performance during the training session. The assumption that raters worked independently was based upon the monitoring of rater performance during the training session. The assumption of rater agreement was based upon training and comparisons of rater recordings of driver performance during the training program.

Subjects were randomly selected from school districts in the greater Lansing area. In the study design, where subjects are not a variable of analysis, the presumption of randomness is not important. The random selection of subjects from several selected programs is included only to reduce the possibility of obtaining a subject population which is not essentially normal.

Definitions of Terms Used

Anchor Points

Anchor points are the extremes in driving behaviors characteristic of satisfactory or unsatisfactory behavior patterns within each Specific Performance Objective Test Sites (SPOTS).

Among Raters

Among raters is the agreement of all raters observing and recording the driver performance of the same subject at different times.

Between Raters

Between raters is the agreement between a pair of raters observing and recording the driver performance of the same subject at the same time.

Direction Control

Direction control is "the driver's coordination of steering and turning maneuvers with speed and timing of steering adjustments."[1]

Driver Behavior Elements

Driving behavior elements are defined as driving behaviors occurring sequentially or simultaneously in response to traffic situations and driving task requirements; i.e., searching, adjusting velocities, accelerating, decelerating and turning in proper time relationships.[2]

---

[1]Nolan, R. O., Vanosdall, F. E., and Smith, D. L., et. al., Driver Performance Research, Final Report, Vol. II, Guide for Training Observer/Raters in the Driver Performance Measurement Procedure. Prepared for National Highway Traffic Safety Administration, Contract FH-11-7627, Michigan State University, Department of Psychology, and Highway Traffic Safety Center, Feb. 1973, p. vi.

[2]Ibid, p. viii.

Driver Performance

Driver performance is that driving performance stipulated by the in-car performance objectives and displayed by the subject driver over a specific route.

Dual Control Vehicle

A dual control vehicle is one which contains an additional brake control mounted for convenient use by the front seat rater.

General Observation Area

The general observation area is "that portion of the route, lying between intensive observation areas, in which the rater is observing the vehicle and driver in relationship to general vehicular placement and maneuvering with respect to other traffic and manmade laws. This area also incorporates the recording area."[3]

In-Car Performance

In-car performance is the performance required of and/or displayed by a driver while preparing to operate and while operating a vehicle in a real-world setting.

Instrument

An instrument is "a set of procedures by means of which an observer can record and categorize the behavior

---

[3]Ibid, p. 49, 273.

or a subject or hazards."[4]

Intensive Observation Area

The intensive observation area is "an area or portion of the route where driver behavior is observed intensively in relation to traffic situations and required driving tasks."[5]

Interrater

See Between Raters.

Narrative

A narrative is a written summary of the driver's performance that can serve as additional documentation regarding the adequacy of the test. It may also provide insight into what the rater is seeing in terms of driver performance. It may also serve to clarify differences between raters in recording driver performances.

Objectivity

Objectivity is the recording, by a rater, of only those behaviors actually observed.

Observational Measure

An observational measure "is a procedure for using an observational record to assign scores to each of the

---

[4]Rowley, Glenn L., American Educational Research Journal, Winter 1976, Vol. 13, No. 1, pp. 51-59.

[5]Nolan, op. cit., pp. 43, 273.

subjects of observation; each score so assigned being assumed to reflect some characteristic of behavior of that subject."[6]

Observer

An observer is synonymous with a rater.

Overt Driving Behaviors

Overt driving behaviors are "behaviors such as head, eye and hand movements that are readily observable physical movements displayed by the subject driver."[7]

Program Effectiveness

Program effectiveness is the extent to which a driver education program of instruction produces a desired effect as determined by a subject displaying performances stipulated by performance objectives.

Rater Mirror

The rater mirror is an extra rearview mirror, mounted by the rater with the aid of suction cups, to assist the rater in identifying driver behavior.

Raters

Raters are those persons who observed and recorded the driving performances of subjects.

---

[6] Rowley, op. cit.

[7] Nolan, op. cit., p. 69.

Recording Area

"The recording area is an area of the driving test route where the recording of observations is carried out."[8]

Reliability

Reliability is the stability, consistency and accuracy with which an instrument measures whatever it does measure.[9]

Run

Run is "one complete circuit of the driving test route."[10]

Search

Search is "an observable behavior in which the driver looks systematically for possible sources of traffic information."[11]

Speed Control

Speed control is "the use of the accelerator or brake to accelerate or slow the vehicle to fit the traffic

---

[8]Nolan, op. cit., p. viii.

[9]Borg, Walter R., Gall, Meredith D., Educational Research: An Introduction (New York: David McKay Co., Inc., 1974), p. 142.

[10]Nolan, op. cit.

[11]Nolan, op. cit., p. viii.

and driving task requirements."[12]

Subjects

Subjects are the drivers having recently successfully completed their program of instruction who were randomly selected from a list of driver education students in the Lansing and East Lansing area.

Time 1 and Time 2

Time 1 and Time 2 refer to a run or negotiation of the same route by the same subject, with different pairs of raters, at different times.

Training Program

The training program is a training program designed to train raters in the use of the instrument. The training program also familiarized them with the route and what it yielded. The training program also consisted of methods used to trigger and observe driver behavior at proper time and space intervals.

Trigger Directions

Trigger directions are directions given at particular locations along the route which initiate a response by the driver to particular driving situations and tasks. These directions also alert the raters to begin intensive observation.

---

[12] Nolan, op. cit.

Validity

Validity of an instrument or test is whether the instrument or test actually measures what it is designed to measure.[13]

Organization of the Remaining Chapters

Chapter II contains summaries of the literature chosen for review. Some studies dealing with plans for investigating driver education, methods of observing and recording driver behavior, driving task analysis and driver education objectives, observational techniques, methods and instrument design were selected for review and reporting.

Presented in Chapter III are (1) route and instrument design, (2) development and administration of a training program for raters, (3) method of obtaining subjects and (4) the collection and analysis of data.

Chapter IV contains the findings; and Chapter V presents the summary and conclusions.

---

[13] Borg, op. cit., p. 135

Chapter II

REVIEW OF LITERATURE

During 1975, the Michigan Department of Education began developing minimal performance objectives to measure the cognitive and psychomotor skills considered basic to any Michigan driver education program. These objectives were finalized in 1976. Appropriate test items were developed for the classroom objectives and pilot tested. A total of sixty objectives were chosen for testing. Each objective was measured by five items with the pass level for each objective set at 80% correct responses. The test results for approximately 100,000 students were analyzed. The findings indicated only thirteen objectives were attained at the 80% correct response level.

This investigation was an integral part of the Michigan Driver Education Evaluation Project. The study dealt with the development of an objective-referenced in-car performance measurement, the development of a route which would yield the necessary opportunities to observe student attainment of the objectives, and the determination of instrument and rater reliability. The project was the first attempt to conduct any systematic research regarding the effectiveness of driver education programs in Michigan.

To allow a determination of the various driver education programs' strengths and weaknesses to be made, objective-referenced tests, rather than norm-referenced tests, were selected. It was determined by project staff that a more immediate and plausible approach to investigating the effectiveness of driver education programs in Michigan, than accident and violation records, was to measure driver education students' attainment of the in-car performance objectives. It was also determined by the project staff that this dynamic approach, rather than an attempt to use the criteria of accidents and violations, would determine the effectiveness of content internalization. An examination of driver behaviors of this nature was more desirable in terms of design, control, expediency and observability.

There have been numerous studies concerned with measuring driver performance and the effectiveness of driver education. While many studies were reviewed and considered, only those having relevance to this study were chosen for reporting. Several studies were selected because they dealt with plans for investigating driver education. Some studies were selected because they contained methods of observing and measuring driver performance. Other studies contained driving tasks and driver education objectives. Some reviews involved studies which described observational methods, techniques and instrument design.

Plans for Investigating Driver Education

The American University Study, by Lybrand, et. al. (1968), supported the approach taken by the Department of Education by expressing the need for accident measures to be carefully qualified when used as driving performance criteria. The study reiterated that accidents are rare phenomena and not a stable characteristic of driver behavior. The study also pointed out that accidents must include a valid measure of exposure if they are to be used as a measure of driver proficiency. The report suggested that driver education programs may be evaluated in terms of their general objectives, enabling objectives or terminal objectives. In summary, Lybrand said, "In order to evaluate proposed driver education programs, there must be a set of instructional objectives derived logically from an adequate description of driving performance and defined in terms of intended behavioral outcomes."[14]

Goldstein (1975) pointed out some observations about accidents and violations of interest. He stated, "Highway accidents have multiple causes and are rare events. Increasing violations may not necessarily increase accidents." Goldstein also pointed out that there appeared to be a long

---

[14] Lybrand, William A., Carlson, Glenn H., Cleary, Patricia A., and Bower, Boyd H. A Study on Evaluation of Driver Education, pp. 210. Report National Highway Safety Bureau, 1968.

delay in driver education students obtaining a license after completion of their course.[15]

William Cole (1976) expressed his concern for the need of performance-based driver education. He stated that, "driver and traffic safety education is performance-oriented, traditionally taught through a sequence of standardized cognitive, affective, and psychomotor learning experiences based on minimum fixed time standards." Cole also made a case against comparing students' performance against their peers. This approach would essentially be a norm-referenced approach to determining effectiveness of driver education. He pointed out that enabling objectives specify what the driver education students should be able to do at the completion of their driver education course. He referred to these enabling objectives as immediate criteria. In his opinion, criterion-referenced tests uniformly applied are preferable to norm-referenced tests. He also stated that there was need for an intermediate criterion that is operationally feasible and statistically reliable if it is used to measure "real-world" driving performance.[16]

---

[15] Goldstein, Leon G., "Rejoinder to Peck and Jones' Reply." Journal of Traffic Safety Education, Vol. XXIII, No. 1, October, 1975, pp. 15 and 17.

[16] Cole, William M. "The Case for Performance-Based Driver and Traffic Safety Education." Journal of Traffic Safety Education, April 1976, Vol. XXIII, No. 3, pp. 9-10.

Warren Quensel (1976) stated that, "evaluation is essentially a process of determining to what extent the program objectives are actually being realized." His approach, however, was to evaluate the ultimate criterion of safe driving by using a questionnaire and self-reporting system to determine accident involvement. He did acknowledge the difficulty in obtaining accurate information from existing state records.[17]

The New York University Report by the Center for Safety (1968) expressed concern for the limited value of studies that determine the effectiveness of driver education programs using the accident criterion. The report stated that, "if we seek to evaluate driver education in terms of accident reduction, we are confronted with so many variables that we become enmeshed in an endless chain of proof." The study encouraged the consideration of short-term, intermediate and long-term criteria. It recommended that driving performance be measured via simulator, road test and self-rated driving knowledge and driving attitudes.

The report also stated there was good reason to consider favorably evaluative techniques concerned with more or less immediate learning as a result of driver education programs. This approach would include driving task analysis in terms of expected behavioral outcomes of

---

[17] Quensel, Warren P. "How to Measure Program Effectiveness." Journal of Traffic Safety Education, April 1976, Vol. XXIII, No. 3, pp. 6.

instruction, development of tests and instrumentation to provide relatively objective ratings of student performance, and a practical design for estimating relative cost effectiveness of varying instructional programs.[18]

Teal, et. al. (1968) pointed out that the general purpose of their research was to develop a concrete plan or plans for evaluating the effectiveness of current or proposed driver education programs. They reported that present methods offered very little insight into the quality of the programs. For a short-term approach, they recommended an evaluative criterion instrument used by a visiting team of teachers at each school. They also recommended a comparative evaluative study, among the various states, for a long-term approach.[19]

Methods of Observing and Measuring Driver Performance

It is noted that, "The Institute for Educational Development Report by Kennedy and Chapman (1968) was critical of driver performance studies where driver performance

---

[18] New York University. Driver Education and Training -- Plans for Evaluating the Effectiveness of Programs, pp. 95. Report No. PH180-473, under Contract FH-11-6560. Washington, D.C.: National Highway Safety Bureau, 1968.

[19] Teal, Gilbert E., Truesdale, Sheridan L., and Fabrizio, Ralph A. Driver Education and Training, pp. 211. Report No. B2D 68-575, Dunlap and Associates, Inc., under Contract FH-11-6559. Washington, D.C.: National Highway Safety Bureau, May 1968.

variables were not derived from, nor validated against, real-world driving situations." The general concensus was performance criteria must be derived from behavior expected of drivers in the real-world. The group stated that since other means were absent, "the best that can be done at this time is to pool the judgment of experts, using what evidence is available in constructing a systematic set of hypotheses about relevant variables and how to measure them." The group also expressed concern for the measurement techniques used. It was felt that checklists could be useful, but considerable care should be taken in setting up the test situation, in defining the scoring basis, and in training the raters.[20]

Quensel, reporting on an in-car evaluation instrument developed by the staff at Illinois State University, stated that, "the best is comprehensive in nature but does not include assessment of basic control skills, procedures for maneuvers, visual habits, identification habits, or the evaluation of hazard." On the same page, he listed the basic response categories. The categories contained the elements of search, speed control, direction control and timing. The assessment of the very things he claimed not to assess appeared contradictory. His criteria for selecting

---

[20]Kennedy, J. L. Driver Education and Training Project, pp. 92. Report No. PH 180 472, under Contract FH-11-6561. Washington, D.C.: National Highway Safety Bureau, 1968.

a route included dynamic situations. Likewise, the concept of drawing inferences from observable behavior was included. He stated the importance of selecting a programmed route, yet apparently 10 of his 35 situations were either not programmed or they were highly atypical. He also reported the need for a reliable scoring system, yet he reported that he could not guarantee that comparable situations would occur for all subjects.[21]

Ulaner et. al. (1952) reported a more reliable and meaningful measure of safe and effective vehicle operation. He explored the assessment of driver behavior through observation by and collective judgments of supervisors and associates of Army drivers. Originally eleven experimental scales were devised. The four that were finally selected were: (1) near accidents; (2) reaction to sudden change; (3) effect of temper on driving; and (4) knowledge of own limitations. The supervisors and peer raters selected 21 items, from a list of 105 driving habits, that they felt they could reliably rate. After review by a panel of experts, these items were reduced to 15.[22]

---

[21] Quensel, Warren P. "An In-Car Evaluation Instrument." Journal of Traffic Safety Education, January 1976, Vol. XXIII, No. 2, pp. 15-16.

[22] Uhlaner, J. E. Development of Criteria for Safe Motor-Vehicle Operation. Highway Research Board Bulletin 60, pp. 36-43. Washington, D.C.: Highway Research Board, 1952.

The "Driver Performance Measurement (DPM) Study" at Michigan State University by Forbes et. al. (1973) was a unique approach to measuring driver performance. Many of the approaches, problems, techniques and concepts used in DPM were helpful in designing this study.

Driver Performance Measurement is a reliable method for research in vehicles on the highway. This procedure for measuring driver performance was intended to be used instead of accident data. It dealt with a wide range of driving behaviors determined to be suitable or unsuitable depending upon the interaction of patterns of behavior with dynamic traffic situations. The evaluation of the behavior pattern was that pattern's reflection upon the increase or decrease of the potential hazard in the situation. Alghouth the "DPM" staff considered the most important test component to be the behavioral pattern, the basic elements of search, speed control and direction control were considered adequate for the present study.

The project staff on DPM reviewed the Human Resource Research Organization (HumRRO) task analysis and found 92.5% of applicable "critical" and "very critical" items were covered. The project staff also performed on-site observations of driver behavior to obtain a "real-world" task analysis of driver behavior.

The route was standardized and arranged in progression with regards to degree of difficulty. An initial warm-up period was provided at the beginning of the route. The six Behavioral Environmental Traffic Situational Sequences (BETSS) selected included urban and rural, two-lane highway, two-lane street, four-lane street, and freeway driving. Controlled and uncontrolled intersections, as well as various types of lane changes, were also included. This was done in an effort to sample a wide range of driving tasks determined by project staff to be important for safe and efficient driving.

In order to assist the observer/rater in determining whether or not the driver's behavior was suitable or unsuitable, the DPM rating form incorporated anchor points. Expected suitable behavior was provided for each behavior pattern to be rated. The exact opposite of expected suitable behaviors was listed for the unsuitable end of the continuum. The unsuitable and suitable behaviors thus provided the anchor points for rating the behavior patterns of the driver. After rating the behavior pattern of the driver as suitable or unsuitable, in relationship to the dynamic traffic situation, the rater then rated the suitability or unsuitability of the elements of driving behavior.

The first attempt at rating driver behavior elements involved seven behavior elements. After pilot observations and ratings, the project staff determined that rating this number of elements was too difficult, especially for the front seat rater. The driving elements to be rated were reduced to: (1) search, (2) speed control, and (3) direction control. Timing was actually a fourth element of driver behavior rated but was considered to be an element that influenced the suitability or unsuitability of the other three elements rather than standing alone.

The route was divided into observation zones and recording zones. The observation zone where the driver's behavior to be recorded was observed was labeled the intensive observation zone. The zone in which the recording of the driver's behavior took place was labeled general observation zone.

Specific standardized directions were identified as the directions that marked the exact time and location that the rater began intensive observation of the driver's behavior. Areas that marked a logical conclusion of the driver's behaviors were identified. These areas marked the end of the intensive observation area and beginning of the general observation area. At this time and location, the observer/rater recalled the behavior just observed and recorded it. It is noted that the raters were encouraged to make marginal notes regarding the driver's performance to assist them in recalling the driver's behaviors.

It should be pointed out that in addition to a training program for raters, the design of the route, incorporation of observation zones and recording zones, and the reduction of the number of driver behavior elements to be recorded contributed to achievement of high agreement between raters using DPM methodology. The reliability estimates for between-raters in the same run were .876 and .946 for behavior pattern and element scores, respectively. For between-raters on different runs, the reliability estimates were .833 and .941 for behavior pattern and element scores, respectively. The training program for DPM raters consisted of 120 hours of training in the concepts and field application of the DPM procedure.[23]

"The Michigan Road Test Evaluation" study by Vanosdall et. al. (1977) incorporated the basic concepts of DPM in measuring driver performance of driver license applicants. Traffic "sequences" and "segments" replaced the "BETSS" and "SubBETSS" used in DPM.

Perhaps the change of most interest was the reduction of the number of environmental situations. The original DPM project incorporated six "BETSS", whereas the Michigan Road Test Project reduced this number to four. The time of the route was reduced from approximately 45 minutes to approximately 20 minutes. The training program for the raters was reduced from 3 weeks to 2 weeks. The final report of this study pointed out that the raters felt they could have benefited from training much sooner if they had been exposed to the test route earlier in the training

---

[23]Forbes, T. W., Nolan, R. O., Schmidt, F. L., et. al. Driver Performance Measurement Research Final Report, Vol. 1, pp. 173. Technical Report, under Contract FH-11-7627. Washington, D.C.: National Highway Traffic Safety Administration, 1973.

program.

The reliability coefficients for raters scoring 288 subjects was reported at .60. The project staff felt this perhaps could have been improved by increasing the number of sequences and early route utilization during the training program for the raters.[24]

The USC on-road performance test by Jones (1977) provided some insight into intensive observation and recording of driver behavior. However, three raters were used, with each rating different tasks or behaviors. This approach was determined to be impractical for the present study due to the desire to develop an instrument that would have the potential to be used by driver education instructors in their courses.[25]

When summarizing earlier studies, Forbes (1950) made his readers aware of some very noticeable factors regarding the observation of drivers. The more experienced drivers picks up minor cues that enable them to anticipate hazards for which novice drivers are apparently not aware. He also pointed out the difference in search behavior of

---

[24] Vanosdall, F. E., et. al. Michigan Road Test Evaluation Study, Final Report, Vol. III. Prepared for National Highway Traffic Safety Administration, under contract MDL-75-002B, Michigan State University, Department of Psychology, Highway Traffic Safety Center, Nov. 1977.

[25] Jones, Margaret Hubbard. Measuring the Outcomes of Driver Training: The USC On-Road Performance Test. Presented at the Transportation Research Board, January 25, 1977.

the experienced versus the novice driver. He suggested that observing the total picture of driver performance, rather than isolated items listed on a checklist, will yield more consistent ratings among observers.[26]

Driver Tasks and Driver Education Objectives

Fine, et. al. (1965), while developing a criterion for driver behavior, called attention to the usefulness of unobtrusive measurement in accident research. They strongly encouraged the use of actual field experiments since such experiments appeared to be more definitive.[27]

The HumRRO staff (1970) developed a comprehensive inventory of the behaviors involved in operating an automobile. One of the primary reasons for developing the task descriptions was to identify a set of driving performances to be used as terminal objectives for driver education courses. Another purpose of the task analysis was to serve as a basis for designing a driver performance test to evaluate the effectiveness of driver education programs.[28]

---

[26]Forbes, T. W. Street and Highway Traffic: Handbook of Applied Psychology. Editors: Fryer and Henry: Rinehard, Vol. 1, 1950, pp. 325-335.

[27]Fine, Jerome L., Malfetti, James L., and Schoben, Edward J., Jr. The Development of a Criterion for Driver Behavior, pp. 43. New York: Columbia University, 1965.

[28]McKnight, J., and Hunt, A. G. Driver Education Task Analysis, Vol. I, Nov., 1970.

The HumRRO staff developed instructional objectives based on the task descriptions. This was to assist in developing instructional programs and the evaluation of those programs.

A driving situations' test was developed. It was intended to evaluate the student's ability to deal with a range of situations that occur in "real-world" driving. The test was designed to be conducted on the road in ordinary traffic. The test was comprised of (1) a list of planned and unplanned driving situations; (2) a checklist of observations for each situation and a format for recording responses; and (3) a set of performance standards.

The report stated that the test was not standardized, but if the test were approximately 30 minutes in duration, the number of responses recorded would be sufficient enough to obtain reliable results. The HumRRO staff pointed out that it was important to plan in advance the route and the observations to be made. Situations should be listed in the sequence in which they occur.

The use of normative data to evaluate driver education students' performance was believed to be inappropriate and of no value. That driver education courses should provide specified minimum standards of qualifications was a major premise. Since the test was used to determine the feasibility of administration, reliability and validity statistics were not computed. A point of interest brought up in the report was that observers were often not able to observe and record all the situations due to the rate of occurrence and spacing of the planned situations. This would indicate a need for separate observational and recording periods.[29]

Observation Methods, Techniques and Design

McGlade (1960) developed an experimental road test. He based the checklist on the information he gathered from forty-six licensing agencies. It reported a relatively high test-retest reliability (r=.77) when used with students who had completed driver education. He used two pairs of raters and achieved interrater reliabilities of (r=.93 and r=.88). The test primarily dealt with the selected skills of braking, parking, right and left turns, lane changes, traffic controls and intersections. The test method evaluated the skills individually rather than looking at a sequence of driver behaviors.[30]

Quenault (1968), in describing a method of systematic observation of driver behavior, pointed out some methods that could be very helpful in observing driver behavior. For example, the use of dual raters was a desirable

---

[29] McKnight, J., and Hunt, A. G. Driver Education Task Analysis, Volumes III and IV, March 1971.

[30] McGlade, Francis Stanley. An Evaluation of the Road Test Phase of the Driver Licensing Examination of the Various States: An Investigation of Current Road Tests and Testing Procedures, and the Development of a Valid and Reliable Road Test Based on Driver Implications, dissertation, pp. 250. New York: New York University, 1960.

technique, but he did not report on interrater reliability. The idea of intensive observation and objectivity was brought up in the study. The use of a rater mirror to assist in observing driver search behavior was mentioned. Other important concepts mentioned were memorization of the route and how to give instructions.[31]

Medley and Mitzel (1963) noted that the observer should not be required to rate behaviors on a quantitative scale, but rather the ratings should be qualitative judgments when possible. The ideal classification task would only involve whether or not the proper behavior was displayed. The simpler the task, the more likely it would be done correctly. It was their conclusion the simplest judgment of whether the behavior occurred or not was best.[32] The authors pointed out that, "selecting behaviors to be observed is done by identifying a limited range of behaviors relevant to the study and constructing items to be used by the observers."[33]

---

[31]Quenault, S. W. Development of the Method of Systematic Observation of Driver Behavior, pp. 50. RRL Report LR 213, Crowthorne, Berks. (Gt. Brit.): Road Research Laboratory, 1968.

[32]Medley, D. M., and Mitzel, H. "Measuring Classroom Behavior by Systematic Observation." In N. L. Cage (Ed.) Handbook of Research on Teaching. Chicago: Rand-McNally, 1963, pp. 251.

[33]Ibid, pp. 251-253.

They summarize an observational technique as, "an observational technique in which an observer records relevant aspects of classroom behaviors as (or within a negligible time limit after) they occur, with a minimum of quantification intervening between the observation of a behavior and the recording of it. Typically, behaviors are recorded in the form of tallies, checks or other marks which code them into predefined categories and yield information about which behaviors occurred or how often they occurred, during the period of observation."[34]

A point of interest was a statement by Medley and Mitzel regarding teacher effectiveness:

> "Since it may be assumed that whatever effect a teacher has on pupils must result from his behaviors, it is only necessary to identify the crucial behaviors, record them, and score them properly to measure effectiveness in process."[35]

The report by Boyd and DeVault (1966) dealt with observational techniques, as well as collecting and recording observational data. The observational techniques reported dealt with participant and nonparticipant observers. The participant observer would be the presence of an observer for the purpose of scientific investigation. By participating in a common natural setting, the observer

---

[34] Ibid.

[35] Ibid.

gathers better data. Most of the research using observational techniques involved the nonparticipant observer. This type of observation frequently utilized mechanical devices for observing and recording behavior. The report pointed out two distinct disadvantages to nonparticipant observation:

1. The increase in cost due to needed hardware.
2. The effect of environmental change on behavior.

The authors pointed out that structured or unstructured observations may be used. The inability of any one observer to see and record all behavior that was displayed and the inability to identify distortions and inadequacies seemed to favor the use of structured observations. They also pointed out that accuracy of recall of information and feedback may effect rater agreement.[36]

Herbert and Attridge (1975) developed a guide for users of observation systems and manuals:

> "A set of thirty-three criteria were identified and sorted into three main types: identifying, validity, and practicality criteria. Identifying criteria enable users to select the correct instrument for their purposes. Validity criteria, which include criteria pertaining to the degree of inference, context, reliability and validity, relate to the accuracy with which the instrument represents the observed events.

---

[36] Boyd, Robert E., and DeVault, M. Vere. "The Observation and Recording of Behavior," Review of Educational Research, 36(5) 1966, pp. 529-551.

> Practicality criteria provide information about the ease of administration and dissemination of results."[37]

This unique guide was very useful in designing the instrument to be used for recording driver behavior. It was also helpful in clarifying procedures that were involved in an observational study of this nature. The criteria developed by Herbert and Attridge, together with the included examples, were very helpful in developing a training program for the raters.

Summary

In this chapter a portion of the literature reviewed and deemed relevant to this study was reported. Summaries of the literature selected for reporting dealt with studies relating to plans for investigating driver education; methods of observing and recording driver behavior; driving task analysis and driver education objectives; and observational techniques, methods and instrument design.

The literature indicated that a standard route should be developed. The route should contain a warm-up section, intensive observation zones and recording zones. The route should provide a range of driving tasks and contain a variety of real-world traffic situations. The

---

[37]Herbert, S. D., and Attridge, C. "A Guide for Developers and Users of Observational Systems and Manuals," American Educational Research Journal, 1975, 12, pp. 1-20.

literature suggested that a route specific rating form, incorporating the use of anchor points to rate behavior patterns of the driver, be developed. The literature also indicated that procedures for using the route and the rating form be developed and used to train raters.

There were several important issues in the literature regarding the observation and recording of behaviors. It was suggested that an examination of driver behaviors stipulated by performance objectives, rather than a norm-referenced approach, was desirable in terms of design, control and observability. The literature indicated that inferences, regarding judgments and decisions, may be drawn from observed behaviors. It was suggested that qualitative judgments of behaviors be used instead of a quantitative scale.

Chapter III

THE METHOD OF PROCEDURE

The primary objective of this study was the development of an instrument to measure the in-car performance of Michigan driver education students and the estimation of the instrument's reliability. An additional concern of the study was the amount of agreement between pairs of trained raters.

This chapter contains the methods of procedure by which the study was conducted. Included are (1) route and instrument design, (2) development and administration of training program for raters, (3) subjects, (4) delimitations, (5) null hypotheses and (6) statistical analysis.

Route and Instrument Design

There were several driving performance tests available for use. However, this project was concerned with the ultimate goal of measuring the effectiveness of Michigan driver education programs. If this practical concern of the project staff was to receive attention, efforts were needed to devise a measure that was responsive and generalizable to the content of Michigan driver education programs that would be evaluated in the future. A practical approach to the development of such a measure seemed to be the development of objective-referenced criteria, using the in-car performance objectives developed for Michigan driver education programs by the Michigan Department of Education.[38]

Route Design

Practical considerations of previous efforts. At this stage of the project, efforts had been made by the Department of Education to develop an instrument and a route. The instrument consisted of 28 typed pages. The format of the instrument was basically that of a checklist requiring the rater to check "Yes" or "No" in response to observations of the driver's performance. The route incorporated approximately 23 miles of roadway and required approximately 50 minutes of driving time.

After reviewing the preliminary instrument and route design, the project staff decided further evaluation and revision was necessary. A panel of experts, consisting of project staff and Michigan State University Highway Traffic Safety Center staff, was formed to assess the project's needs and to examine the preliminary route and instrument design.

The panel was comprised of experts in the areas of human factors' research, driver education, driver licensing, psychology and evaluation. The panel and project staff reviewed the instrument and the route individually and

---

[38] Michigan Department of Education, Driver Education Performance Objectives, June 1976.

collectively by both reading the materials and driving the route. The combination of expertise, experience and group discussions complemented a practical approach to analyzing the preliminary design and content of the instrument and route. The combined efforts of the panel of experts provided the basis for the development of an instrument and route that was manageable and comprehensive. This approach met both the practical and research concerns of the project staff.

Determining route content. Following the recommendations of the panel of experts, the project staff's next task was the evaluation of the in-car performance objectives. The project staff directed their review of the objectives to the determination of which objectives were critical to safe operation of a vehicle in a real-world traffic environment, as well as to those objectives which were atypical, extremely hazardous, or were logical prerequisites to the attainment of other objectives.

It was the decision of project staff that certain objectives, such as those covering parking, turnarounds, backing, and entering and leaving the car, were procedurally and manipulatively oriented. These objectives were deleted from the set to be evaluated as they were determined not to be critical for safe operation of the vehicle. These objectives could be evaluated by a separate instrument, using an off-street area, if desired in future projects.

The project staff also decided that objectives requiring responses to emergency vehicles, school buses, and passing maneuvers were conditions that offered a very low probability of occurrence for each subject being evaluated. On the basis of this observation, the decision was made to eliminate these objectives from the evaluation.

The objectives requiring subjects to demonstrate the procedures for off-road recovery and operating the vehicle without power assistance were considered too hazardous to include in this evaluation. The safety and liability issues involved warranted their omission from the study.

It was agreed by project staff that various objectives and procedures, such as placing the gear selector in park or neutral and turning the key to start before starting the engine, were obvious prerequisites to terminal objectives. However, provisions for measuring these objectives were provided for in the vehicle familiarization section of the evaluation by a Yes/No checklist.

The route was then reviewed and analyzed to determine what tasks were required for proper negotiation. These task requirements were then matched with the task requirements of the performance objectives. A determination was then made on whether the route yielded the necessary situations to evaluate the performance specified by the objectives chosen for evaluation. Portions of the route that yielded no situations, or at least no new situations, were identified. The route was then modified to eliminate or reduce the length of nonproductive segments of the route.

The route was reviewed again by project staff. Consideration was given to areas that might cause unwanted delays in negotiating the route. After locations such as railroad crossings, parade routes, and construction zones were identified, the route was further revised.

Selecting a starting point. Having determined the approximate beginning and end of the test route, it was necessary to locate an area that could serve as the origin and final destination of the road test during data collection. Ideally, such an area would provide off-street parking, a facility to shelter staff and subjects from the weather and restroom facilities. The facility should be close to the test route and provide relative ease of entry to the beginning of the route.

With the aid of a Lansing-East Lansing map, such an area was located. The Red Cedar School was identified as a potential location. After an inspection of the facilitiy, it was determined that the facility was nearly ideal in meeting the specifications. Project staff from the Michigan Department of Education made the necessary inquiries and requests, and permission was obtained to use the Red Cedar School.

With the route's origin now determined, an adequate warm-up section could be added to the route. This would allow the subject an opportunity to become more familiar with the operating features of the vehicle, the presence of the raters and the manner in which directions would be given. It also provided an opportunity for the raters to create a more relaxed atmosphere for the subject and to identify extreme deficiencies in the subject's ability to operate the vehicle.

A practical return and closure of the test route could also now be determined. The portion of the route incorporated for the return to the facility was analyzed to determine if that portion of the route contained any situations that would yield opportunities and requirements for measuring performance specified in the performance objectives.

Identifying intensive and general observation areas. The route was further reviewed to identify the exact areas where an objective or combination of objectives would be evaluated. The task was to identify a logical beginning and ending of the driver behaviors required to complete the driving task and attain the objective specified. The route between designated points where intermittant testing began and ended was referred to as *intensive observation areas*. Areas between the intensive observation areas were identified or created. These were areas where driver performance was not being observed for the purpose of recording the driver's behavior. These areas were labeled *general observation areas* and provided the time and distance along the test route for the rater to complete a record of the driver performance that had just been observed in the intensive observation area of the route. This provided for a comprehensive route yet one which was manageable by raters. The route covered approximately 11 miles and required an average driving time of 28 minutes.

Developing standardized directions. The next component in the route design was the development of a standardized set of directions. These would be used by the raters to direct drivers over the test route. The directions had to be clear and direct. They would have to be specific and use common terminology as much as possible. The wording was designed to initiate a response on the part of the driver, but not to change the driver's behavior from what would normally be displayed during the execution of various driving tasks. Landmarks and traffic signals were utilized to clarify the directions. At times hand gestures were used to complement the directions. The wording, timing and location along the route were to be standardized components of the directions. The front seat rater was to give the directions. If the driver asked for the directions to be repeated or indicated the directions were not clear, the rater was to repeat the directions.

The directions were piloted by administering the directions to volunteers as they drove over the test route. On the basis of the information received from the pilot test, the directions were revised. The directions were then presented to the panel of experts used in the review of the route and instrument content. However, the directions were not finalized until the raters had an opportunity to use them during the training program. The standardized directions can be found in Appendix A.

Identifying and controlling abort situations. While reviewing the route and pilot testing the directions, the writer noticed that there were several locations along the route that were conducive to either a driver or a traffic abort situation. If the driver deviated from the prescribed route, due to driver error or traffic interference, the planned observations and the recording of driver behavior would be interrupted. Either alternate test segments to the route would have to be designed, or an alternate route returning the driver to the original test route or coaching to avoid the abort would have to be used. Due to the nature of the potential abort locations, the project staff decided to use the latter two options. For those locations that provided the capability of easy return to the route, that technique would be used. For those abort situations that would cause considerable increase in time of returning to the route, coaching was determined to be the most practical means of avoiding the abort. An example would be to coach a driver to position his/her vehicle into a certain lane ahead of time to insure that the abort situation was avoided. The coaching was to be done while the driver was in a general observation area. Although the coaching would alter the driver's behavior, it would occur in an area where the driver's behavior was not being observed for the purpose of recording. However, if the rater had to intervene while the driver was in an intensive observation area, that maneuver or behavior would be recorded as unsatisfactory.

Instrument Design and Format

Designing the rating form. To assist in reliably scoring driver performance, a form was needed that would permit the rater to recall the driver's performance and record it quickly and accurately. Each page of the rating form was labeled a LOPE. (LOPE is an acronym for Location Of Performance Evaluation). Each LOPE is an intensive observation area and is comprised of test segments or traffic situations referred to as SPOTS. (SPOTS is an acronym for Specific Performance Objective Test Site). The rating form was designed to record four scores for each test segment (SPOTS). The four scores were pattern, search, speed control and direction control. The exception to recording four scores would be when a performance objective stipulated timing as an additional performance to be scored. The Driver Performance Measure (DPM) and Michigan Road Test (MRT) rating forms were used as models for the present rating form and provided the following:

1. A summary of specific behaviors, stipulated by the performance objectives, listed for each SPOTS. These anchor points assisted the raters in recalling the satisfactory/unsatisfactory pattern performance for the SPOTS.

2. Satisfactory and unsatisfactory rating spaces for each behavior pattern and element for each SPOTS.

3. Space for qualitative notes or abbreviations that would assist the rater in recalling what had happened.

4. A logical progression for scoring as LOPE's and SPOTS followed the exact route.

5. A separate page for scoring each LOPE.

Immediately after scoring each LOPE, the page was to be turned and the rater prepared to enter the next LOPE. The rating form can be found in Appendix B.

Procedure for driving performance. The method of scoring driver performance required the trained rater to observe intensively the driver's performance throughout the LOPE. Upon the driver's completion of the LOPE, the rater was to record immediately the driver's performance as "satisfactory" or "unsatisfactory" for each SPOTS as stipulated by the performance objective. The record of performance was completed by first scoring the overall pattern of performance for the SPOTS and then, secondly, scoring the element behaviors. Each score required the rater to know the driving task and the range of satisfactory behavior required to complete the task as stipulated by the performance objective. Observation of the driver's performance for the purpose of scoring occurred only during the LOPE. Recording of the observed driver performance occurred during the time and distance between the LOPE's. After recording was completed and while still between LOPE's, directions were given to the next portion of the route. This method of scoring driver performance in the driver education road test was the same as used in the DPM and the MRT.

Dividing the rating form into two formats. The first format was labeled vehicle familiarization. The subject was asked, by the front seat rater, to identify the gauges and devices by pointing them out or touching them, when possible. The front seat rater then proceeded to state verbally the information gauges, starting and control devices, and safety devices that appeared in order on the rating form.

Beside each gauge or device on the rating form was a Yes/No column. The raters then placed a check by the appropriate column to indicate the subject's response. This rating was done shortly after the subject entered the vehicle and before the subject prepared to move the car from the parking area.

Other items found on the vehicle familiarization form were listed under the headings of pre-ignition control tasks, starting the engine, putting the car into motion, stopping the vehicle and securing the vehicle. The rater made no verbal request of the subject to perform these tasks other than to give directions to initiate the exit from and return to the parking lot. After giving the directions necessary to initiate these behaviors, the raters independently observed and recorded the subjects' performance.

The second format of the rating form consisted of recordings of the subject's driving performance as the subject responded to various driving tasks while negotiating the route. Each page of this format was identified as a LOPE. Each LOPE was identified by a Roman numeral. The numerical order of the LOPE's represented the sequential progression of the portions of the route where driver behavior was intensively observed and recorded. A sample of the MDE rating form can be found in Figure 1.

Each LOPE was subdivided into segments referred to as SPOTS. Directly under the heading SPOTS, numbers, representing the performance objectives being tested, were listed. Beside the SPOTS was listed the specific location on the route where the observation of the performance specified by the performance objective would occur. Immediately following the specific location where the driver behaviors were to be observed was listed the range of expected driver behaviors. The listed driver behaviors served as anchor points and represented the extreme unsatisfactory and satisfactory behaviors on a continuum. The satisfactory behaviors were representative of the behaviors specified in the performance objectives. The extreme unsatisfactory behaviors were determined by using the opposite behaviors specified by the performance objectives.

Previous research, project staff and the panel of experts agreed that all observable driver behaviors could be recorded under three basic elements of the driving task.

Figure 1

Michigan Driver Education Evaluation Project
Driver Performance Rating Form

Subject ______________ Program ________________
Rater ______________
Date ______________
Run No. ______________

LOPE II	Specific Performance Objective Test Site		Search	Speed Control	Direction Control
				Performance On	
Spots 3.1	Michigan Avenue westbound		U____S	U____S	U____S
	U	S			
	Does not check mirrors, fails to signal right, does not check blind spot, does not change to right lane, changes lanes abruptly causing traffic to slow or swerve, does not adjust lane position or speed, fails to cancel directional signal.	Checks mirror, signals right, checks blind spot, changes to right lane, blends smoothly with traffic, adjusts lane position and speed, cancels directional signal.			
Spots 3.3A	Michigan Avenue and right turn onto Homer		U____S	U____S	U____S
	U	S			
	Fails to check left and rear traffic, fails to signal right, fails to reduce speed, starts turn early, turns into lane #1 or 2, recovers by palming or shuffling wheel, does not adjust speed to flow.	Checks traffic especially left and rear, signals right, reduces speed, turns into lane #3, recovers using hand over hand, adjusts speed to match flow.			

These elements were search, speed control and direction control. It was also agreed that a fourth element existed but that it was a determining factor in the satisfactory or unsatisfactory rating of the three major elements. This element was timing.

After giving a direction to the driver, the raters began intensive observation of the driver's behaviors during the LOPE. Upon completion of the designated LOPE, the driver entered a general observation area. While in the general observation area, the driver's behavior was not intensively observed; rather, the raters were recalling and recording the driver's performances of the previous LOPE. The elements of search, speed control and direction control were rated as unsatisfactory or satisfactory based upon the driver's compliance with the satisfactory anchor points of each of the SPOTS within the completed LOPE.

Some of the objectives specified timing as a criterion for satisfactory performance. For those objectives, a _T_ was printed immediately under the unsatisfactory symbol of the rating form. If the element of performance was unsatisfactory due to timing, both the _U_ and _T_ were marked. If the driver performed a particular element satisfactorily, the performance was rated satisfactory by marking the _S_.

Development and Administration of Training Program for Raters

Selection of Raters

The study involved six raters. Five of the individuals selected as raters had completed the educational requirements for certification to teach driver education in Michigan. Of these, four had recently completed a course at Michigan State University consisting of supervised practical teaching experience in driver education, and the other had been teaching driver education in the public schools for approximately 12 years. The sixth rater was a member of the Michigan Department of Education who had participated in the rater training program.

The design called for two pairs of raters to be used in the data collection phase of the study. These two pairs of raters were counter-balanced, and they rotated front and rear seat positions on two successive drives with a subject. The four most compatible raters, excluding the rater from the Department of Education, as determined from practice runs, were to be used in the data collection phase. The remaining rater, along with the individual from the Michigan Department of Education, formed a pair of alternate raters and were assigned an alternate subject. Used in this manner, the additional raters, subject and vehicle were always available to serve as a backup to cover vehicle malfunction or any absenteeism of subjects and raters but were not used in the data analysis. The data from the practice runs indicated that there was no obvious rater incompatibility; therefore, the four raters used in the data collection and analysis were randomly assigned to pairs.

Training Program

The six individuals selected as raters were administered a four-day training program. The major elements of the training effort were route-specific. Stating the directions verbatim at precise locations along the route, knowing when to observe and when to record, and recognizing where expected behaviors would occur were critical to objective and accurate recording of the observed driving behaviors. This necessitated a major block of time for practical work in a vehicle on the route.

Other methods for presenting route-specific training included the use of 35 mm slides of the route and overhead transparencies of the route and rating form. The color slides represented the route components and were arranged in progression. The slides were taken from inside an automobile from the front seat passenger's side. This was done to better represent the front seat rater's view. In addition to representing the progression of the route, the slides depicted the nature of dynamic traffic patterns and represented a range of satisfactory driver behavior for each SPOTS. The overhead transparencies were used either alone or as complements to the slides of the route. When used in combination, the slides and transparencies were incorporated in a split-screen technique.

In addition to the route-specific training, the raters were exposed to new terminology related to the study. Techniques for giving directions and for responding to the subjects' questions were presented and demonstrated. Techniques for observing and recording driver behavior, avoiding abort situations and maintaining the safety of the driver and vehicle were also presented and demonstrated. Practical application of these training components was attained by actual practice on the route using practice subjects. The raters were paired. Then they directed the subject over the route and observed and recorded the driver's performance. The writer monitored these practice runs and discussed them with the raters.

During the training program, Dr. Robert O. Nolan and Mr. Fred Vanosdall discussed and demonstrated writing narratives of driver performance and objectivity in recording driver performance, respectively. The narratives provided a qualitative summary or explanation of the driver's performance. The session on the need for objectivity in recording driver performance provided the rationale for drawing inferences and making judgments based only upon observed behavior.

There was one training session during the program that involved the use of films. The two films used were from the Aetna Driver Simulator series and were entitled <u>IPDE</u> and <u>Separate and Compromise</u>. Although these films were not related to the test route, they did offer the opportunity for the raters to view the same driving tasks and driver behaviors in a controlled environment. Only the introductory portion of the film, which incorporated the use of a model driver, was used. The raters were asked to observe driver behaviors and to recall and list them when the projector was stopped. This was done in short segments with the only narrative being verbal directions for the model driver. The list of behaviors was then shared and discussed. The film was re-run during the discussion to clarify the observable behaviors. The exercise provided an example of objectivity, reliability and intensive observation.

The training material was delivered primarily through a lecture method. These lectures were supplemented with audio-visual presentations, when appropriate. Practical applications of the training program were conducted during field exercises that involved practice ratings of subjects on the test route. The outline of the training program can be found in Appendix C.

Selection of Practice Subjects

While the training program for raters was being conducted, driver education students from Lansing Catholic Central High School were completing their driver education course. These students were asked to volunteer to participate in the training program by driving a prescribed route while a pair of raters observed and recorded their driving performances. The subjects were told that they would be asked to drive the route twice.

Volunteers were asked to complete the name, address and phone number sections on a parent permission form, have their parents sign the form and return the form to the writer. The students were told that they would be called ahead of their scheduled session to finalize arrangements.

The students were called two days prior to their scheduled participation in the training program to confirm their attendance. On the day they were scheduled to drive, the students were picked up at their home and returned when their driving session was completed.

The students who participated as practice subjects during the training program had just completed a four-phase driver education program. The students received the classroom phase of their course at Lansing Catholic Central High School; the simulation, driving range and on-street phases were conducted at Michigan State University as part of a driver education teacher preparation course. In this course university students performed the practical teaching responsibilities under the supervision of MSU Highway Traffic Safety Center staff. All of the practice subjects had passed their driver education course, but none had yet received his or her driver's license.

Evaluation and Feedback

To assist in determining the effectiveness of the training program and the progress of the raters, several techniques of evaluating each rater's progress were used. Evaluation of rater performance and feedback were concentrated on comprehension of the route and route components and on rater agreement on driver performance ratings.

To evaluate the raters' comprehension of the route and route components, the raters were requested to trace the test route on a map. The raters were then asked to identify the beginning and ending of each LOPE on the map. The raters were asked to state verbatim the directions for the route while viewing slides of the route. The raters were also required to state the driving task and satisfactory driving behaviors for each SPOTS, while viewing the slides of the route.

To evaluate rater agreement on recorded behaviors, the raters were required to view a portion of film representing a demonstration driver and record the driver's behaviors. The recorded behaviors by the raters were then compared for agreement. The film was re-run to reinforce standard observations and recordings.

During early practice runs on the route, the ratings were reviewed and discussed after each LOPE. While monitoring practice runs with practice subjects, the writer also rated the subjects and compared these ratings to those of the raters. Rating forms were monitored after each practice run. In addition to comparing the ratings of paired raters, the writer checked the rating forms for omitted data and margin notes and asked the raters to recall their observations when there was disagreement between raters or when data was missing from the rating form.

Subjects

During the month of July, the Michigan Department of Education identified seven school districts in the Lansing-East Lansing area having driver education programs that would be completed prior to the scheduled data collection phase of this study.

Eighty students from the seven school districts were randomly selected as potential subjects for the driver performance test. Although only 45 subjects would be required for the study, the extra students were necessary in the event some students had time conflicts, did not receive permission to participate or failed to return the permission forms. Letters were sent to the 80 students' parents or guardians requesting permission for the students to participate. From the 45 students required for the study, 30 would be evaluated by two pairs of raters. The additional 15 students would be assigned to an alternate pair of raters and used as alternate subjects if some of the original 30 students could not participate.

The 80 subjects, randomly selected for participation in the data collection phase of the study, were from the

Dansville, Haslett, Lansing, Mason, Okemos, Waverly and Williamston school districts. The students had completed a driver education program just prior to the data collection phase of the project. All subjects selected had passed a driver education course but were not yet licensed to drive.

The subjects chosen for participation came from a combination of classroom and on-street content to a combination of classroom, simulation, range and on-street educational programs. The type of program was not a consideration of this study. The raters, however, were not told what type of program the subjects had completed. This was done to avoid the possibility of rater bias in regards to their opinions as to which programs may produce better drivers.

Delimitations

Based upon the characteristics of the subject population discussed in the preceding section, generalization of the findings is limited to:

1. Students who have successfully completed a driver education program which used Michigan Driver Education Performance Objectives and who were not yet licensed to drive.

2. Students in the age range of 14 to 18.

3. Students whose socio-economic background is consistent with that found in the greater Lansing area, which includes inner city, rural and suburban populations.

Hypotheses

Study Analysis

There were five major hypotheses investigated in this study. The focus of the first three hypotheses concerned the consistency of test scores across items and test runs. The following were the specific null hypotheses tested:

1. Research Hypothesis: There would be differences in difficulty among the items in the driver performance test.

   H_o: Item difficulty will not have a systematic effect on driver performance scores.

2. Research Hypothesis: Run administrations would not effect driver performance scores.

   H_o: Test runs will not have an effect on driver performance scores.

3. Research Hypothesis: The subjects' driving performance scores would not vary according to items interacting with the time of test administration (i.e. between run one and run two).

   H_o: Driver performance scores will not vary according to items interacting with run administrations.

The focus of the fourth hypothesis was on the internal consistency of the test. This hypothesis considered the relationship between true scores and observed scores.[39] The hypothesis also took into account the effects of time and individual items.

---

[39] Glass, Gene V. and Stanley, Julian C. Statistical Methods in Education and Psychology. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.

4. Research Hypothesis: A positive relationship would exist between true driver performance scores and observed driver performance scores.

   H_o: No relationship exists between true driver performance scores and observed driver performance scores.

The focus of the fifth hypothesis was interrater reliability, or the agreement between pairs of raters.

5. Research Hypothesis: A positive relationship would exist between raters on measures of sum, search, speed control, direction control, familiarization and signs.

   H_o: No relationship exists between raters on measures of sum, search, speed control, direction control, familiarization and signs.

Statistical Analysis

The purpose of the study was to determine the reliability of an observational measure designed to evaluate in-car performance of Michigan driver education students. It was important, therefore, to determine the various reliability characteristics of the in-car performance test procedures. Topics to be discussed in this section are:

a. counterbalanced design.

b. test for null hypotheses 1-3 using ANOVA.

c. test for hypothesis 4 - reliability coefficients and significance.

d. test for hypothesis 5 - reliability coefficients and significance.

Counterbalanced Design

There were thirty study subjects making two test runs. There were four raters evaluating the two runs. A counterbalanced design was used for efficiency in the analysis by compensating for external influences such as rater bias in terms of rating run 1 for a subject versus run 2, as well as assuring that no rater would be consistently paired with another rater. Since four raters were used, a repeated pattern of rater assignment, counterbalanced at run 1 and run 2 for each group of six subjects, would cover all possible rater pairings. For example, raters A and B tested subject 1 during run 1 and subject 6 during run 2. (The counterbalanced design is shown in Figure 2). This design also provided for control of contemporary history, maturation processes, measuring instruments, statistical regression, experimental mortality and interaction of selection and maturation.

Test of Null Hypotheses 1-3

The first three null hypotheses for the study were tested using analysis of variance. This test was chosen for its ability to test for separate effects of two or more independent variables and the interaction effects of those variables. In this study, the variables were items, runs and subjects.

Sources of variation were determined using the Millman-Glass Rules of Thumb (Ref). The relevant ANOVA Table is provided in Figure 3.

For hypotheses 1-3 the following F-ratios were used.

Figure 2

Counterbalanced Design - Repeated Pattern of Rater Assignment

Subjects	Run 1 Raters ABCD	Run 2 Raters ABCD
	Items for each rater for each run	
1	AB	CD
2	A C	B D
3	A D	BC
4	BC	A D
5	B D	A C
6	CD	AB

Repeated pattern of rater assignment for each group of 6 subjects.

Figure 3

ANOVA Table

Sources	SS	df	
Between Subjects	SS_s	29	$MS_s=SS_{s/29}$
Within Subjects			
Items	SS_I	I-1	$MS_I={}^{SS}I/(I-1)$
Runs	SS_t	1	$MS_t={}^{SS}t/1$
Items x Runs	SS_{ti}	(I-1)1	$MS_{IT}={}^{SS}IT/(I-1)1$
Items x S_s	SS_{Is}	29(I-1)	$MS={}^{SS}IS/29(I-1)$
Runs x S_s	SS_{ts}	29	$MS={}^{SS}IS/29$
Items x Runs x S_s	SS_{ITs}	29(I-1)	$MS={}^{SS}ITS/29(I-1)$

1. Differences in item difficulties

$$F = \frac{\text{MS Items}}{\text{MS Items x Subjects}}$$

2. Differences in average performance over runs

$$F = \frac{\text{MS Runs}}{\text{MS Runs x Subjects}}$$

3. Items x Runs interaction; i.e., is pattern of performance on the items consistent at Runs 1 and 2

$$F = \frac{\text{MS Items x Runs}}{\text{MS Items x Runs x Subjects}}$$

Test of Null Hypothesis 4

Reliability coefficients are affected by the assumptions one makes regarding the sources of variation built into the study. The assumptions used here treated subjects as a random variable and items and runs as fixed variables. These assumptions lead to a liberal interpretation of the data, hence, a higher expected reliability coefficient. One could, however, treat subjects and items as random, while leaving runs fixed, or subjects, items and runs as random. Both sets of assumptions lead to progressively more conservative estimates of reliability since they will account for a smaller true score.

This study reported the estimates of reliability for all three methods to provide for the possible range of reliability estimates. Each of the formulas for the reliability coefficients had the form of $\frac{\text{True score variance}}{\text{Observed score variance}}$.

Method I Subjects Random-Items fixed - Runs fixed

$$\frac{\text{MS Subject-MIXSX Runs}}{\text{MS Subjects}}$$

Method II Subjects Random Items Random Runs Fixed

$$\frac{\text{MS Subjects-MS Items x Subjects}}{\text{MS Subjects}}$$

Method III Items Random Subjects Random
Runs Random

$$\frac{\text{MS Subjects} + \text{MS}_{\text{IX SXR}} - \text{MS}_{\text{IXS-SXR}}}{\text{MS Subjects}}$$

The findings were considered significant if the F ratio was beyond that expected at alpha = .01.

Test of Null Hypothesis 5

Since the assumption of normality and equality of variance was made, a parametric statistic was needed. The analysis used continuous data. The Pearson Product Moment correlation was used to determine the correlation coefficient. The findings were considered significant if the coefficient was greater than that expected at alpha = .01.

Summary

This chapter contained the methods of procedure by which the study was conducted. Included were (1) route and instrument design, (2) development and administration of training program for raters, (3) subjects, (4) delimitations, (5) null hypotheses and (6) statistical analysis. Remaining are Chapter IV, in which the findings of the study will be presented, and Chapter V, which will report a summary of the results and conclusions.

Chapter IV

FINDINGS OF THE STUDY

The primary purpose of this study was to determine the reliability characteristics of an in-car performance test. The test was developed to measure driver education students' attainment of the in-car performance objectives set by the State of Michigan. More specifically, the study was concerned with whether or not there was variability in test performance (measured by variance in item difficulty), whether raters could consistently rate a driver's performance within the same run and whether one run affected student performance on a second run.

The remainder of this chapter will present the results of the reliability study within each of the null hypotheses set out in Chapter III. The first section covers item difficulty and score stability. This section presents data concerning the first three null hypotheses. The next section, reliability of the in-car performance measure, presents the comprehensive reliability coefficients, taking into account multiple raters, runs, items and subjects. The last section, interrater reliability, presents correlations between pairs of raters.

Item Difficulty and Rating Stability

The test was expected to show a range in difficulty for the various component parts in order to determine variance in driver performance. This was necessary in order to ensure that a subject's score on one component of the test was not necessarily a predictor of the total test and that subjects truly vary in ability to perform the in-car objectives. The study also intended to demonstrate that there was stability between performances from one run of the test to the next. Significant differences between scores on run one and run two could reflect a learning effect. Rater bias is not an issue since no rater made back-to-back runs with the same subject.

Following are the three null hypotheses tested to determine the consistency and stability of ratings for items and runs of the two test administrations.

1. Item difficulty will not have an effect on driver performance scores.
2. Run administrations will not have an effect on driver performance scores.
3. Driver performance scores will not vary according to items interacting with run administrations.

All three null hypotheses were tested with ANOVA. The ANOVA Table can be found in Table 1.

Following are the F ratios for tests of the first

Table 1

ANOVA Table

Relationships Between Student Performance, Item Difficulty and Run Administrations

Source of Variance	SS	DF	MS	F	S
Items	SS_I	I-1	$MS_s=SS_{s/29}$	41.75	.01
Runs	SS_t	1	$MS_I=SS_{I/(I-1)}$	3.36	NS
Subjects	SS_s	29	$MS_s=SS_{s/29}$		
ItemsxSubjects					
ItemsxRuns				1.03	NS
ItemsxSubjectsxRuns					

three null hypotheses.

1. Item difficulty will not have an effect on driver performance scores.

$$F = \frac{\text{MS Items}}{\text{MS ItemsxSubjects}} = 41.75$$

The F ratio was significant beyond alpha .01. Therefore, the null hypothesis was rejected. This finding suggested that there was a difference in item difficulty and that the items were discriminating.

2. Run administrations will not have an effect on driver performance scores.

$$F = \frac{\text{MS Runs}}{\text{MS RunsxSubjects}} = 3.36$$

The finding was not significant. Therefore, the null hypothesis was not rejected. It did not appear that back-to-back runs made a difference in driver performance. This result was encouraging in that performance appeared to be stable over runs, thereby suggesting that the student may not necessarily "learn" by taking the test when there is no feedback given after the run.

3. Driver performance scores will not vary according to items interacting with run administrations.

$$F = \frac{\text{MS ItemsxRuns}}{\text{MS ItemsxRunsxSubjects}} = 1.03$$

There was some concern that while no overall item or run differences might occur, there might be an interaction between the two. If this were supported, then one could assume that some items in the test were sensitive to learning effects. However, the finding ($F = 1.03$) was not significant. Therefore, the null hypothesis was not rejected. This suggested that there was no significant interaction between test items and run administrations.

Internal Consistency

Hypothesis four stated that no positive relationship exists between true driver performance scores and observed driver performance scores. There are several ways to approach this relationship, and depending upon the assumptions one makes, the resulting correlation coefficient is affected.

If subjects are treated as a random variable, then it can be suggested that subjects are a random sample from the population of subjects that could be tested with the instrument. Items and runs were treated as fixed variables, considering the items as the only items of interest to measure the objectives and the runs as the only two runs of interest. Therefore, the resulting reliability coefficient is generalizable to other subjects from the same population, but only to those items included in the test and the two runs administered. This particular method of analysis is perhaps the most conservative because of its limitations on generalizability.

For the purposes of this study, two additional methods of analysis were performed. Each method is progressively more generalizable. In addition to considering subjects as a random sample of other subjects from the same population, items may be considered a random sample from those items measuring the same objectives.

The third method of analysis is the most liberal treatment of the variables. With this method one can generalize to all subjects, all items and all testing times from the respective populations of subjects, items and testing times.

Following are the results of the three methods of analysis. The formula applied for the analysis is taken from the Millman-Glass Rules of Thumb for the analysis of variance. The reliability coefficient has the following form:

$$\frac{\text{True Score Variance}}{\text{Observed Score Variance}}$$

Method I. Subjects Random Items Fixed Runs Fixed

$$\frac{MS_s - MS_{IxSxR}}{MS_s}$$

$$\frac{6.803728-.261101}{6.083728} = .957$$

There was a positive relationship between true and observed driver performance scores. Therefore, the null hypothesis was rejected. The reliability coefficient for the overall test was very high and clearly acceptable.

Method II. Subjects Random Items Random Runs Fixed

$$\frac{MS_s - MS_{IxS}}{MS_s}$$

$$\frac{6.083728-.379508}{6.083728} = .937$$

The reliability coefficient remains very high and acceptable.

Method III. Subjects, Items and Runs All Random

$$\frac{MS_s + MS_{IxSxR} - MS_{IxS-SxR}}{MS_s}$$

$$\frac{6.083728+.261101-.379508-1.518399}{6.083728} = .730$$

Interrater Reliabilities

Hypothesis five concerned the interrater correlations between ratings of Sum, Drive, Search, Speed Control, Direction Control, Familiarization and Signs.

In addition to the reliability coefficients of the test, rater agreement was determined for each of the test components. Because the same two raters were not always paired, the rater agreement was determined for pair one (irrespective of individuals) on run one and pair two (irrespective of individuals) on run two. The correlations are presented in Table 2.

The correlations showed a high degree of agreement on the overall (SUM) test ratings. Rater agreements on the other components were also high ranging between .49 and .83. This suggested that regardless of pairing or front and back

Table 2

Rater Agreement

Components	Pair One	Pair Two
Drive	.80	.80
Search	.67	.79
Speed Control	.72	.51
Direction Control	.75	.50
Familiarization	.83	.52
Signs	.49	.54
Sum	.86	.83

seat positions, raters of similar background, who were administered the same training objectives, could be expected to use this instrument with a high degree of consistency. There was a positive relationship between ratings; therefore, the null hypothesis was rejected.

Summary

In this chapter was found the analysis of the reliability characteristics of the Michigan Driver Education Test. The analysis addressed the areas of item difficulty and stability, internal consistency and interrater reliabilities. The following chapter contains a summary of the results and conclusions.

Chapter V

SUMMARY AND CONCLUSIONS

Summary

This study dealt with the development of an observational measure for evaluating in-car performance of Michigan driver education students and the determination of the measure's reliability characteristics. The development of the observational measure involved the development of a standard route, a route-specific instrument and procedures for scoring.

Reliability and validity had to be determined if the instrument was to be used to assist in the evaluation of driver education program effectiveness. To be responsive to the concerns for the reliability characteristics of the instrument, it was necessary to determine item difficulty and rating stability, internal consistency and interrater reliabilities. This study determined that the instrument was reliable and could be used consistently by trained raters.

A validation study, using the Michigan State University Driver Performance Measurement criterion, was conducted at a later date. This study determined that the instrument was valid.[40]

Statement of the Problem

No instrument existed that was designed to measure the Department of Education's in-car performance objectives. An instrument of this nature was needed by the Department of Education to determine the effectiveness of driver education programs in the state.

The purpose of this study was to develop an observational measure to evaluate in-car performance and to determine the reliability characteristics of the observational measure. It was important to ascertain which of the observational measures provided by the instrument were reliable, which were not, and under what conditions. In keeping with this concern, the study addressed the following research hypotheses:

1. That there would be differences in difficulty among the items in the driver performance test.
2. That run administrations would not affect driver performance scores.
3. That the subjects' driving performance scores would not vary according to items interacting with the time of test administration (i.e. between run one and run two).

---

[40]Michigan Department of Education. Michigan's Driver Education Evaluation Project. Lansing: The Department, 1978.

4. That a positive relationship would exist between true driver performance scores and observed driver performance scores.
5. That a positive relationship would exist between raters on measures of sum, search, speed control, direction control, familiarization and signs.

Methods of Procedure

A concern of the project staff was the development and evaluation of a driver performance measure that was responsive and generalizable to the content of the programs that would be evaluated in the future. An approach to the development of such a measure was the development of objective-referenced criteria, using the in-car performance objectives developed for Michigan driver education programs by the Michigan Department of Education.

A panel of experts, consisting of project staff and Michigan State University Highway Traffic Safety Center staff, was formed to assess the project's needs and to examine the preliminary route and instrument design. The combined efforts of the panel of experts provided the basis for the development of an instrument and route that was manageable and comprehensive.

The initial task of the project staff was the evaluation of the in-car performance objectives. The project staff directed their review to the determination of which objectives were critical to safe operation of a vehicle in a real-world traffic environment, as well as to those objectives which were atypical, extremely hazardous, or were logical prerequisites to the attainment of other objectives.

An integral part of the study process was the design of a test route that would yield the necessary situations to observe and record student driving performance as stipulated by the Michigan Department of Education's in-car performance objectives. The second component of the process was the design of an instrument that was concise, thorough and definitive enough to be easily manageable by the raters observing and recording driver performance. The third component was the design and implementation of a training program for the raters who observed and recorded the driving performances. The final component was the development and implementation of a counterbalanced design for rater and subject assignment during the data collection phase of the study, and the statistical treatment of the data to determine the reliability characteristics of the driver performance test.

Major Findings

The first hypothesis, "Item difficulty will not have an effect on driver performance scores," was not rejected. The F ratio was significant beyond alpha .01, suggesting that there was a difference in item difficulty.

The second hypothesis, "Run administrations will not have an effect on driver performance scores," was not rejected. The finding was not significant, suggesting that the performance appeared to be stable over time.

The third hypothesis, "Driver performance scores will not vary according to items interacting with run administrations," was not rejected. The finding was not significant and suggested that there was no significant interaction between test items and run administration.

The fourth hypothesis, "No positive relationship exists between true driver performance scores and observed driver performance scores," was rejected. Three methods of analysis, each progressing in the assumption of randomness of the variables, were employed. The correlation coefficients were .957, .937 and .730, respectively.

The fifth, and final, hypothesis, "No relationship exists between raters on measures of sum, search, speed control, direction control, familiarization and signs," was rejected. The interrater reliability for pairs of raters ranged from .49 to .83 on test components. The sum, or overall test, had a reliability coefficient of .86 for pair one and .83 for pair two.

Conclusions

The results of this study indicated that the driver performance measurement test, developed to measure the in-car performance of Michigan driver education students as stipulated by the Michigan Department of Education's in-car performance objectives, was a reliable test. It had a range of difficulty in the items, was internally consistent, and had consistency and stability of ratings across two runs of administration. The interrater reliabilities appeared to be more than adequate, meaning that the test was dependable under the conditions of the rater training program. Whereas a reliable and valid test did not exist to measure driver performance, as stipulated by Michigan's driver education performance objectives, prior to this study, one now exists.

Recommendations

Based upon the results of the test developed, the issue is to adopt a more widespread use of the test. If the Department of Education should choose to modify or change the performance objectives for the in-car phase of driver education, then the test would need additional development. If the test is put to use, then consideration must be given to the efficient training of people to use the test. Based upon the results of the study, the following recommendations are made.

1. It is recommended that the Department of Education use this test procedure to measure the attainment of in-car performance objectives for successful completion of driver education courses.
2. It is recommended that the Department of

Education use this test procedure to measure program effectiveness.

3. It is also recommended that this test procedure be used to measure the effectiveness of methods and materials and delivery formats of various driver education programs.

The above recommendations should seriously be considered as a means of addressing the issues of accountability, cost effectiveness and teacher merit.

4. It is recommended that the Department of Education use this test procedure as the criteria for evaluating competency-based driver education programs and the students' attainment of the in-car performance objectives for competency-based programs.

In addition to the test being a valid criterion for pre and post evaluation of student performance, the test, with the use of well-designed feedback, could be used as a teaching aid.

5. It is recommended that the Department of Education not permit this test procedure to be used to evaluate student performance or program effectiveness, without ensuring that the raters have been adequately trained to develop a route or use the instrument.

If this test is used by untrained persons or by persons trained under conditions other than those set forth in this study, the test cannot be considered dependable.

Recommendations for Further Research

Due to the results of the study and its potential for widespread application, the following recommendations for further research are made:

1. It is recommended that the research be replicated with a population from different programs, different geographical locations and from different socioeconomic backgrounds.
2. It is recommended that a study be undertaken to formalize the training program administered in this study for raters. It is also recommended that the formalized training program be pilot tested before final adoption.
3. It is recommended that an off-street testing procedure be developed to accommodate the performance objectives considered too hazardous or occurring too infrequently to be measured on the street. These types of objectives were not measured in this study.
4. It is also recommended that a time series study be conducted to determine the instrument's potential for predictive validity in regards to predicting what type of accident or violation a subject might experience at some future point in time.

5. It is recommended that an effort be undertaken to formalize route-development procedures, as this is an important part of the test.

Discussion

During the preliminary stages of the study, it became apparent that there was a need for an objective-referenced test to evaluate the effectiveness of driver education. As development efforts proceeded, the need for realistic and clearly defined performance objectives became apparent. It was obvious that the initial effort put into the development of these objectives would affect the quality of an objective-referenced test. It is not only crucial that the objectives be stated in clearly observable behavioral terms, but also that they provide for the conditions under which the objective will be taught and the behavior observed. It is also important to specify what degree of attainment is satisfactory.

During the route development phase of the study, it became apparent that driver education teachers and evaluators need to know how to effectively develop a standardized route that will yield the opportunity to evaluate and record student driving performance. In order for the route to yield the opportunity for reliable evaluation, the development and coordination of driving tasks, performance objectives and traffic situations must be done by on-sight observation and verification by the developer, rather than an armchair concensus by so-called experts.

During the training program for raters, a combination of lectures, audio visuals and field exercises were used. The slides of the route and expected driver behaviors were photographed from the front seat passenger side of the vehicle. This photographic perspective allowed the situation to be displayed from the front seat rater's perspective. This technique appeared to be effective. If more than one training route or testing route is developed in the future, it is recommended that programmed training materials, incorporating the use of detailed sketches of the route, be used rather than a slide program. This approach would probably be as effective and would definitely be less expensive.

It is the writer's opinion that the sooner the trainees are introduced to the route and the more practice rating they are exposed to, the more effective the training will be. By using a comparison of percentage of agreement, the raters' learning curve or rate of agreement seemed to peak on the second day of the data collection phase of the study. The rater agreement might have peaked sooner if an additional day of training had been conducted or if practice rating had started sooner. Based on this experience, there is a danger in exposing the rating form too soon or to persons who have received no training. The untrained person is likely to perceive the rating form as a recipe or checklist, capable of being used by anyone who has taught driver education.

During the data collection phase of the study, the raters were asked to write a brief narrative of the driver's performance. Although these data were not used in the analysis, they did provide a qualitative explanation of the recorded driver performance. This technique appeared to convince the raters to make definitive marginal notes about the driver's performance on the rating form. This technique would be very useful during feedback sessions with a student. Although feedback was not provided to subjects during the study, it would be a necessity if the instrument were used as a teaching aid.

Now that a valid objective-referenced driver performance test exists, there are implications for further development, research and change for driver education in Michigan. This study has implications for curriculum change, diagnostic and learning effects, teaching methods versus student performance, route parameters and efficiency and psychological functions relating to judgments and decisions related to operating a vehicle.

Now that a valid instrument for measuring program effectiveness in Michigan exists, it is time to affect program changes in Michigan driver education.

APPENDICES

APPENDIX A

DIRECTIONS FOR MDE ROAD TEST

APPENDIX A

Directions for MDE Road Test

1. Turn left and proceed to the end of the street; then turn right.

2. Proceed to the traffic light and turn left.

3. Proceed to the third traffic light and turn left.

4. (Frandor sign) Proceed to next light and turn right.

5. (Mister D sign) Turn left at the second traffic light (Student should remain in lane 2).

6. (Past Howard St. light) Turn right at the next traffic light.

7. Proceed to the third street on the left and turn left; then proceed to the end of the street and turn left.

8. Proceed to the end of the street and turn right (after turn, tell student to be in lane 3).

9. (After the Grand River fork - student in lane 3) Proceed to the third traffic light and turn left.

10. Turn right at the second traffic light; (after completing turn) Proceed to the second light and turn left (Student should be in lane 2).

11. Continue to the second traffic light and turn right.

12. Proceed to the second traffic light and turn left; (after turn) proceed to the next light and turn left (from second lane).

13. Proceed to the first street after the traffic light and turn right; (after the turn) continue to the third street on the left and turn left.

14. Proceed to the first street and turn right; continue to the second traffic light and turn left (lane 2).

15. Proceed ahead and enter the expressway East 496.

16. Exit at the East Lansing/Flint exit; continue to the first traffic light and turn right (take the East Lansing turn off).

17. Proceed straight ahead.

18. (After crossing the bridge) Turn right at the first street on the right; turn right at the next street.

19. Turn left at the second street and continue ahead.

20. (After crossing Larkspur) Turn right at the next street; proceed ahead and return to the parking lot on the right.

APPENDIX B

MICHIGAN DRIVER EDUCATION EVALUATION PROJECT
DRIVER PERFORMANCE RATING FORM

APPENDIX B

MICHIGAN DRIVER EDUCATION EVALUATION PROJECT
Driver Performance Rating Form

VEHICLE FAMILIARIZATION

Subject ______________ Program ______________
Rater ______________
Date ______________
Run No. ______________

Spots		Parking Lot		
1.1	A.	Identify Information Gauges		
	a.	Alternator Light (Gauge)	YES	NO
	b.	Brake System Warning Light	YES	NO
	c.	Fuel Gauge	YES	NO
	d.	Left and Right Turn Light	YES	NO
	e.	Odometer	YES	NO
	f.	Oil-Pressure Warning Light (Gauge)	YES	NO
	g.	Seat Restraint Light	YES	NO
	h.	Speedometer	YES	NO
	i.	Temperature Indicator Light (Gauge)	YES	NO
1.1	B.	Starting and Control Devices		
	a.	Accelerator	YES	NO
	b.	Footbrake	YES	NO
	c.	Gear Shift Selector	YES	NO
	d.	Ignition and Starter Switch	YES	NO
	e.	Park Brake	YES	NO
	f.	Steering Wheel	YES	NO

1.1	C.	Safety Devices		
	a.	Door Locks	YES	NO
	b.	Emergency Flasher Control	YES	NO
	c.	Head Testraints	YES	NO
	d.	Headlight Beam Switch and Indicator	YES	NO
	e.	Heater and Defroster	YES	NO
	f.	Horn	YES	NO
	g.	Light Switch	YES	NO
	h.	Rearview and Sideview Mirrors	YES	NO
	i.	Seatbelt Restraint System	YES	NO
	j.	Sunvisor	YES	NO
	k.	Windshield Wiper and Washer	YES	NO
1.3		Pre-Ignition Control Tasks		
	a.	Enter Vehicle (Checks for traffic as situation requires)	YES	NO
	b.	Places Key in Ignition	YES	NO
	c.	Locks all Doors	YES	NO
	d.	Adjusts Seat to Suitable Position	YES	NO
	e.	Adjusts Head Restraint	YES	NO
	f.	Adjusts Mirrors	YES	NO
	g.	Fastens Safety Restraining Devices	YES	NO
	h.	Makes Sure Park Brake is ON	YES	NO
1.4		Starting the Engine		
	a.	Presses Accelerator and Releases	YES	NO
	b.	Depresses Foot Brake	YES	NO
	c.	Puts Gear Selector in PARK or NEUTRAL	YES	NO
	d.	Turns Key to START and Releases when Engine Starts	YES	NO

1.5		Putting the Car in Motion		
	a.	Depresses Foot Brake	YES	NO
	b.	Selects Proper Gear	YES	NO
	c.	Releases Park Brake	YES	NO
	d.	Checks Mirrors	YES	NO
	e.	Uses Proper Signal	YES	NO
	f.	Checks Blind Spot	YES	NO
	g.	Releases Foot Brake	YES	NO
	h.	Gradually Accelerates into Proper Lane	YES	NO
1.7		Stopping the Vehicle		
	a.	Checks Mirrors	YES	NO
	b.	Positions Car Appropriately	YES	NO
	c.	Releases Accelerator	YES	NO
	d.	Brakes to Smooth Stop	YES	NO
1.8		Securing Vehicle		
	a.	Shifts to PARK gear	YES	NO
	b.	Sets Park Brake On	YES	NO
	c.	Turns Off Ignition	YES	NO
	d.	Removes Key	YES	NO

MICHIGAN DRIVER EDUCATION EVALUATION PROJECT
Driver Performance Rating Form

Subject______________ Program____________
Rater ______________
Date ______________
Run No.______________

			Performance on	
LOPE I	Specific Performance Objective Test Site	Search	Speed Control	Direction Control
Spots 1.6	Harrison Road Northbound approaching Kalamazoo U ______ S Places hand on lower half of wheel, steers with one hand, does not maintain proper lane position, does not adjust speed to conditions, does not systematically search. — Hands on upper half of steering wheel, maintains proper lane position, adjust speed to conditions, and searches systematically.	U____S	U____S	U____S
SPOTS 5.1 5.2 4.2	Harrison Road Northbound to Michigan Avenue U ______ S Does not search all directions. Fails to adjust speed to conditions, fails to maintain lane position, fails to observe traffic signals. — Searches all directions, adjusts speed to conditions, maintains proper lane position, observes traffic signals.	U____S	U____S	U____S
SPOTS 2.1 4.4A I1 4.4B I2	Harrison Road turning left onto Michigan Avenue U ______ S Does not reduce speed, signal left, and check traffic in all directions; positions car too far to the right, fails to yield to oncoming traffic and pedestrians; palms or shuffles wheel when turning or recovering, turns into right lane, fails to adjust speed to flow. — Reduces speed and signals left, checks traffic in all directions, positions car close to center line, yields to traffic and pedestrians, uses hand over hand when turning and recovering, turns into lane #1 or #2, adjusts speed to match flow.	U____S	U____S	U____S

DYP-DYV

Figure 1

Michigan Driver Education Evaluation Project
Driver Performance Rating Form

Subject _______________ Program __________________
Rater _______________
Date _______________
Run No. _______________

LOPE II	Specific Performance Objective Test Site		Search	Performance On: Speed Control	Performance On: Direction Control
Spots 3.1	Michigan Avenue westbound U	 S	U___S	U___S	U___S
	Does not check mirrors, fails to signal right, does not check blind spot, does not change to right lane, changes lanes abruptly causing traffic to slow or swerve, does not adjust lane position or speed, fails to cancel directional signal.	Checks mirror, signals right, checks blind spot, changes to right lane, blends smoothly with traffic, adjusts lane position and speed, cancels directional signal.			
Spots 3.3A	Michigan Avenue and right turn onto Homer U	 S	U___S	U___S	U___S
	Fails to check left and rear traffic, fails to signal right, fails to reduce speed, starts turn early, turns into lane #1 or 2, recovers by palming or shuffling wheel, does not adjust speed to flow.	Checks traffic especially left and rear, signals right, reduces speed, turns into lane #3, recovers using hand over hand, adjusts speed to match flow.			

MICHIGAN DRIVER EDUCATION EVALUATION PROJECT
Driver Performance Rating Form

Subject______________ Program______________
Rater ______________
Date ______________
Run No.______________

LOPE III	Specific Performance Objective Test Site		Search	Speed Control	Direction Control
Spots 3.3C	Homer Street and left on Grand River U	S	U_____S	U_____S	U_____S
	Does not search in all directions, fails to position car in lane #2, fails to signal left, fails to reduce speed or keep wheels straight when stopped, turns into lane #1, recovers by palming or shuffling wheel, fails to adjust speed to flow.	Searches all directions, positions car in lane #2, signals left, reduces speed, keeps wheels straight when stopped, turns into lane #2, #3, or #4, recovers using hand over hand, adjusts speed to flow.			
Spots 3.3B	East Grand River and right onto Foster U	S	U_____S	U_____S	U_____S
	Fails to search in all directions, fails to position car in lane #4, fails to signal or signals left, does not reduce speed, fails to check mirror, starts turn too soon causing right rear tire to strike curb or too late causing vehicle to enter the lane of oncoming traffic, does not recover by using hand over hand steering.	Searches in all directions, positions car in lane #4, gives right signal, reduces speed, checks mirrors, starts turn when front wheels are opposite point where curb begins to curve, turns into lane #1, recovers using reversed hand over hand, adjusts speed to flow.			

MICHIGAN DRIVER EDUCATION EVALUATION PROJECT
Driver Performance Rating Form

Subject______________ Program______________
Rater ______________
Date ______________
Run No.______________

LOPE IV	Specific Performance Objective Test Site		Performance On: Search	Performance On: Speed Control	Performance On: Direction Control
Spots 5.1 5.2	Foster Street northbound crossing Woodruff U Does not search in all directions, fails to reduce speed as approaching the intersection, fails to maintain lane position.	 S Searches systematically in all directions, reduces speed as approaching intersection, maintains lane position.	U_____S	U_____S	U_____S
Spots 5.1 5.2 5.3	Foster Street turning left onto Hopkins U Does not search in all directions, fails to signal left, fails to maintain proper speed and lane position, turns too fast, turns too soon, or does not maintain control.	 S Searches systematically in all directions, signals left, reduces speed and maintains lane position, accelerates smoothly, begins turn just before front bumper reaches center of intersection, maintains control while recovering to proper lane position.	U_____S	U_____S T	U_____S T
Spots 5.1 5.2 5.3	Crossing Hayford U Does not search all directions, accelerates or maintains speed, searches only after entering intersection, reduces speed only after entering intersection, fails to stay in own lane.	 S Searches all directions before entering intersection, reduces speed as approaching intersection, stays in own lane.	U_____S T	U_____S T	U_____S

MICHIGAN DRIVER EDUCATION EVALUATION PROJECT
Driver Performance Rating Form

Subject______________ Program______________
Rater ______________
Date ______________
Run No.______________

LOPE IV (2)	Specific Performance Objective Test Site		Search	Speed Control	Direction Control
				Performance On	
Spots 5.1 5.2 5.3	Crossing Magnolia U Does not search all directions, accelerates or maintains speed, searches only after entering the intersection, reduces speed only after entering intersection, fails to stay in own lane.	S Searches all directions before entering intersection, reduces speed as approaching intersection, stays in own lane.	U_____S T	U_____S T	U_____S
Spots 5.1 5.2 5.3	Crossing North Fairview U Does not search all directions, accelerates or maintains speed, searches only after entering the intersection, reduces speed only after entering intersection, fails to stay in own lane.	S Searches all directions before entering intersection, reduces speed as approaching intersection, stays in own lane.	U_____S T	U_____S T	U_____S
Spots 5.1 5.2 5.3	Left turn onto Wood Street U Does not signal left, fails to search rear and continuously right and left, stops where visibility is poor, accelerates jerkily, turns too soon or too late, does not use hand over hand, fails to accelerate to flow.	S Signals left, searches rear, searches continuously right and left, stops in position to see traffic right and left, gradually accelerates and starts turn just before reaching center of intersection, uses hand over hand, accelerates to flow.	U_____S T	U_____S T	U_____S T

MICHIGAN DRIVER EDUCATION EVALUATION PROJECT
Driver Performance Rating Form

Subject____________ Program____________
Rater ____________
Date ____________
Run No.____________

LOPE V	Specific Performance Objective Test Site		Search	Speed Control	Direction Control
				Performance On	
Spots 3.3A	Wood Street turning right onto Grand River U	S	U____S	U____S	U____S T
	Does not check traffic thoroughly, does not position car to right side of lane, fails to signal right, fails to stop and keep wheels straight, does not check mirrors, starts turn too soon or too late, does not turn into lane #4, does not use hand over hand to recover, fails to adjust speed to traffic flow.	Checks traffic thoroughly, positions car to right, gives right signal, stops with wheels straight, checks mirrors, starts to turn when front wheels are opposite point where curb begins to curve, turns into lane #4, recovers using hand over hand, adjusts speed to flow.			

MICHIGAN DRIVER EDUCATION EVALUATION PROJECT
Driver Performance Rating Form

Subject____________ Program____________
Rater ____________
Date ____________
Run No.____________

			Performance On		
LOPE VI	Specific Performance Objective Test Site		Search	Speed Control	Direction Control
Spots 3.1 5.3	North Cedar Street (multiple lane change to right) U	S	U____S	U____S T	U____S T
	Does not check mirrors, fails to signal right, does not check blind spot, changes to far right lane in one motion, interferes with traffic, does not remain in own lane and accelerate to flow, fails to cancel signal.	Checks mirrors, gives right signal, checks blind spot, moves into lane #2 and adjusts speed to flow.			
Spots 3.1 5.3	North Cedar Street U	S	U____S	U____S T	U____S T
	Does not check mirrors, fails to signal right, does not check blind spot, interferes with traffic, does not move into lane #3 and adjust speed to flow, does not cancel turn signal.	Checks mirrors, gives right turn signal, checks blind spot, moves into lane #3, adjusts speed and position, cancels turn signal.			
Spots 3.1 5.3	East Ottawa (multiple lane change left) U	S	U____S	U____S T	U____S T
	Does not check mirrors, fails to signal left, fails to check blind spot, interferes with traffic, does not move into lane #3, does not adjust speed and direction, fails to cancel turn signals.	Checks mirrors, signals left, checks blind spot, moves into lane #3, does not interfere with traffic, adjusts position and speed, cancels turn signals.			
	Repeat to lane #2	Repeat to lane #2	U____S	U____S T	U____S T
	Repeat to lane #1	Repeat to lane #1	U____S	U____S T	U____S T

MICHIGAN DRIVER EDUCATION EVALUATION PROJECT
Driver Performance Rating Form

Subject____________ Program____________
Rater ____________
Date ____________
Run No.____________

LOPE VII	Specific Performance Objective Test Site		Performance On: Search	Speed Control	Direction Control
Spots 4.1 4.2b 4.2f 4.4A.I)1 4.4A.II 4.4B.I)1 4.4B.II 4.3a	Townsend crossing Washtenaw U Does not stop before entering intersection, does not search continuously, fails to yield to pedestrians/vehicles, encroaches on other lane when crossing.	S Stops before entering intersection, searches continuously, yields to pedestrians/vehicles, maintains lane position while crossing.	U_____S	U_____S	U_____S DYP-DYV
Spots 4.1 4.2b 4.2f 4.4A.I)1 4.4A.II 4.4B.I)1 4.4B.II 4.3a	Townsend crossing W. Kalamazoo Street U Does not stop before entering intersection, does not search continuously, fails to yield to pedestrians/vehicles, encroaches on other lane when crossing.	S Stops before entering intersection, searches continuously, yields to pedestrians/vehicles, maintains lane position while crossing.	U_____S	U_____S	U_____S DYP-DYV
Spots 4.1 4.2b 4.2f 4.4A.I)1 4.4A.II 4.4B.I)1 4.4B.II 4.3a	Townsend turning left onto W. Lenawee Street U Does not stop before entering intersection, does not search continuously, fails to yield to pedestrians/vehicles, encroaches on or turns into oncoming lane.	S Stops before entering intersection, searches continuously, yields to pedestrians/vehicles, turns into correct lane.	U_____S	U_____S	U_____S DYP-DYV

MICHIGAN DRIVER EDUCATION EVALUATION PROJECT
Driver Performance Rating Form

Subject______________ Program______________
Rater ______________
Date ______________
Run No.______________

LOPE VIII	Specific Performance Objective Test Site		Search	Speed Control	Direction Control
				Performance On	
Spots 4.5 5.1 5.2 5.3	E. Main Street to 496 East U Fails to use acceleration lane, does not signal left, fails to check traffic thoroughly including mirrors and blind spot, does not accelerate to match flow, merges across first lane, fails to center car in lane and adjust to flow quickly, fails to cancel signal.	 S Enters acceleration lane, signals left, checks traffic thoroughly including mirror and blind spot, accelerates to flow, merges into nearest lane, centers car in lane and adjust speed to flow immediately, cancels turn signal.	U____S T	U____S T	U____S T
Spots 4.6 5.1 5.2 5.3	496 East approaching East Lansing/Flint Exit U Fails to position in far right lane, fails to signal right, does not check traffic thoroughly, fails to check mirror, fails to check blind spot, fails to enter deceleration lane early, does not adjust to exit speed.	 S Positions car in far right lane, gives right signal, checks traffic thoroughly including mirror and blind spot, enters deceleration lane and slows to exit speed.	U____S T	U____S T	U____S T

MICHIGAN DRIVER EDUCATION EVALUATION PROJECT
Driver Performance Rating Form

Subject______________ Program______________
Rater______________
Date______________
Run No.______________

Performance On

LOPE IX	Specific Performance Objective Test Site		Search	Speed Control	Direction Control
Spots 4.4A II 4.4B II 5.1 5.2	Daisy eastbound approaching Larkspur Drive U Fails to adjust speed to conditions, fails to search continuously in all directions, encroaches on oncoming lane, fails to yield to pedestrians/traffic.	 S Adjusts speed for conditions, searches continuously in all directions, maintains lane position, yields to pedestrians/vehicles.	U_____S	U_____S DYP-DYV	U_____S
Spots 4.4A.I)2 4.4B.I)2 5.1 5.2 5.3	Crossing Larkspur U Fails to reduce speed, fails to search continuously left/right, does not yield to pedestrians/vehicles, accelerates abruptly.	 S Reduces speed, searches continuously left/right, yields to pedestrians/vehicles, accelerates smoothly.	U_____S T	U_____S T DYP-DYV	U_____S
Spots 4.4A II 4.4B II 5.1 5.2	Daisy eastbound approaching Narcissus U Fails to adjust speed to conditions, fails to search continuously in all directions, encroaches on oncoming lane, fails to yield to pedestrians/vehicles.	 S Adjusts speed for conditions, searches continuously in all directions, maintains lane position, yields to pedestrians/vehicles.	U_____S	U_____S DYP-DYV	U_____S
Spots 4.4A.I)2 4.4B.I)2 5.1 5.2 5.3	Daisy turning right on Narcissus U Does not signal right, maintains speed or accelerates, fails to search continuously left/right, yields to pedestrians/vehicles.	 S Signals right, reduces speed, searches continuously left/right, yields to pedestrians/vehicles.	U_____S T	U_____S T DYP-DYV	U_____S

MICHIGAN DRIVER EDUCATION EVALUATION PROJECT
Driver Performance Rating Form

Subject______________ Program______________
Rater ______________
Date ______________
Run No.______________

LOPE X	Specific Performance Objective Test Site	Performance On Compliance with Michigan Vehicle Code when Encountering:	
Spots	General Observation Over Route		
4.1	Signs	Signs	Comments
	a. Warning	Yes ___ No ___	
	b. Regulatory	Yes ___ No ___	
	c. Service and Guide	Yes ___ No ___	
4.2	Traffic Signals	Traffic Signals	
	f. Traffic Control Signals	Yes ___ No ___	
4.3	Pavement Markings	Pavement Markings	
	a. Center Lines	Yes ___ No ___	
	b. Crosswalk Lines	Yes ___ No ___	
	d. No Passing Zones/Lines	Yes ___ No ___	
	e. Solid Yellow Lines	Yes ___ No ___	
	f. Turn Lanes/Lines	Yes ___ No ___	

APPENDIX C

OUTLINE OF RATER TRAINING PROGRAM

APPENDIX C

Outline of Rater Training Program

Day 1: Thursday, July 28

8:15 a.m. I. Reliability
- A. Definition of reliability
- B. Necessity of identifying the same behaviors

8:30 a.m. II. Observation
- A. Observation of driver behavior rather than the environment
  1. Head
  2. Eyes
  3. Feet
  4. Mirror
  5. Hands

8:45 a.m.
- B. Observation of driver behavior in relationship to the environment
  1. Lane position
  2. Spacial relationships
  3. Traffic
  4. Pavement markings and signals

9:15 a.m.
- C. Intensive observation of driver behaviors
- D. General observation of driver behaviors
- E. Behaviors directly observable
  1. Search - eye and head movement
  2. Speed control - accelerating, decelerating, braking, kinesthetic value
  3. Direction control - hand movement, vehicle alignment, spacial relationships, lane position, signals

9:45 a.m. III. Basic driving functions
- A. Search
- B. Speed control
- C. Direction control
- D. Timing (early or late) - an element that affects all functions

1. Increases or decreases hazards
2. Affects smooth or abrupt steering or acceleration
3. Affects search (ability to gather information)
4. Affects signals (turns, lane changes)
5. Affects crossing, joining and leaving traffic

Break

10:30 a.m. IV. Inferences
- A. Drawing inferences based on observed behavior
- B. Judgments, predictions, reasoning, decisions - must observe driver behavior first

10:45 a.m. V. Lane Numbering
- A. Lanes numbered from left to right
  1. One-way street - begins at far left side of street
  2. Two-way street - begins in first lane to the right of the center line
  3. Divided two-way street - begins in the first lane to the right of median or barrier
- B. Diagrams of lane numbering

11:15 a.m. VI. Aborts
- A. Traffic abort - traffic mix is so dense it is impossible to maneuver. "Return to Route"
- B. Driver abort - missed directions, wrong turn or lane change, drives past entrance or turn. "Score unsatisfactory"
- C. Rater abort - lane directions, safety reasons. "Score satisfactory"
- D. Coaching

11:40 a.m. VII. Practicality
- A. Legality - compliance
- B. Safety - rater responsibility
- C. Rater must distinguish between legal vs. safe behavior

Lunch

1:00 p.m. VIII. Dual Raters
- A. Front seat rater - directions, safety, observation, dual controls
- B. Rear seat rater - observation, safety if necessary

1:15 p.m. IX. Directions
- A. Precise, consistent
- B. Timing - trigger directions (triggers behavior)
- C. Clues - landmarks
- D. Driver - recognition and compliance
- E. Reminder - point out or gesture with and, repeat
- F. Changing or altering driver behavior

1:45 p.m. X. Comments
- A. Marginal notes - words, phrases, abbreviations, symbols Examples - Lt, Ls, RL, Dyp, Dyv, 2 fast
- B. Narratives of driver performance
  1. Dr. Robert O. Nolan, speaker

Break

2:45 p.m. XI. Ride Route (2 vehicles - 3 runs)
- A. Rotate monitoring; rotate rater positions
- B. Directions
- C. Mirror placement
- D. Review directions - rater comments and input on directions

Day 2: Friday, July 29

8:15 a.m. I. Practical Work on Route
- A. Front seat rater - gives directions, observes, list observed behaviors
- B. Use mirrors
- C. Rear seat raters - observe and record
- D. Stop for discussion after each LOPE

Break

10:30 a.m. II. Discussion
- A. Directions
- B. Trigger directions
- C. Behavior - overt and inferred
- D. Reasons for differences

11:00 a.m. III. Definitions

- A. Dynamic Traffic Environment - vehicles, pedestrians, road surfaces, weather, vegetation, movement of vehicle, traffic controls
- B. Driver Interaction - with dynamic traffic environment
- C. Combinations of Behaviors - example: turns often involve combinations of behaviors
- D. Overt Behavior - directly observable or perceived through sensory processes (kinesthetic value)

11:30 a.m. IV. Type of Observations

- A. During intensive observation
- B. During general observation

Lunch

1:00 p.m. V. Audio Visuals

- A. Transparencies for each LOPE
- B. Transparencies for instrument (rating form)
- C. Slides of route
- D. Film - "IPDE" and "Separate and Compromise" (introduction portion only) Exercise: rate model driver as satisfactory or unsatisfactory for each - search, speed control, direction control
- E. Slides of route and instrument simultaneously (use split screen)

Break

2:45 p.m. VI. Practical Work on Route (2 vehicles - 6 raters)

- A. Adult licensed drivers (volunteers)
- B. Raters - give directions, complete rating form, (drive portion only), rotate positions
- C. Discussion after each complete drive

Day 3: Saturday, July 30

8:30 a.m. I. Practice Runs (1 vehicles - 6 raters)

- A. High school subjects
  1. Six subjects in the a.m.
  2. Six subjects in the p.m.
- B. Complete entire rating form

C. Rater pairs (rotate front and rear seat positions after each run)
D. Subjects (rotate cars after each run)
E. Write narrative for the last subject run in the a.m. and the p.m.
F. Monitor rating form completion after each subject

Day 4: Monday, August 1

8:15 a.m. I. Review
- A. General LOPE 10
- B. First recall
- C. Independent recording
- D. Rolling stops, running light, crosswalks (recording)

9:30 a.m. II. Objectivity - speaker, Mr. Fred Vanosdall

Break

10:45 a.m. III. Review
- A. Vehicle familiarization

Lunch

1:00 p.m. IV. Practical Work on Route (3 vehicles - 6 raters)
- A. High school subjects (6 subjects in the p.m.)
- B. Complete entire rating form
- C. Rater pairs (rotate positions after each run)
- D. Subjects (rotate cars after each run)
- E. Write narrative (for the last subject)
- F. Oral review of ratings by each pair of raters

BIBLIOGRAPHY

BIBLIOGRAPHY

Borg, Walter R. and Gall, Meredith D. Educational Research, An Introduction. New York: David McKay Company, Inc. 2 editions, 1971.

________. Educational Research, An Introduction. New York: David McKay Company, Inc., 1974, p. 142.

Boyd, Robert D. and Devault, M. Vere. "The Observation and Recording of Behavior," Review of Educational Research, 36 (5) (1966), 529-551.

Cole, William. "The Case for Performance-Based Driver and Traffic Safety Education." Journal of Traffic Safety Education, April 1976, Vol. XXIII, No. 3.

Cronbach, L. J., Gleser, G. C., Nanda, H., and Rajartnam, N. The Dependability of Behavioral Measurements: Generalizability of Scores and Profiles. New York: Wiley, 1972.

Denton, G. G. "The Effect of Speed Change on Drivers' Judgment." RRRL Report LR 97, pp. 6. Crawthorne, Berks (Gt. Brit.): Road Research Laboratory, 1967.

Ebel, R. L. "Estimation of the Reliability Ratings." Psychometrika, 1951, 16, 407-424.

Edwards, D. S. and Hahn, C. P. Filmed Behaviors as a Criteria for Safe Driving. Report No. AIR-C80-2/70-FR, prepared for National Institutes of Health, 1970.

Fine, Jerome L., Malfetti, James L., and Shoben, Edward J., Jr. "The Development of a Criterion for Driver Behavior," pp. 43. New York: Columbia University, 1965.

Forbes, T. W. "The Normal Automobile Driver as a Traffic Problem." The Journal of General Psychology, 20, 1939, pp. 471-474.

________. Street and Highway Traffic, Handbook of Applied Psychology. Editors: Fryer and Henry; Rinehart, Vol. 1, 1950, pp. 325-335.

Forbes, T. W., Nolan, R. O., Schmidt, F. L., et. al. Driver Performance Measurement Research Final Report, Vol. 1, pp. 173. Technical Report, under contract FH-11-7627. Washington, D.C.: National Highway Traffic Safety Administration, 1973.

Frick, T. and Semmel, M. I. "Observational Records: Observer Agreement and Reliabilities." Paper presented at the 1974 meeting of the American Educational Research Association, Chicago, April 16, 1974.

Glass, Gene V. and Stanley, Julian C. Statistical Methods in Education and Psychology. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.

Goldstein, Leon G., "Rejoinder to Peck and Jones' Reply." Journal of Traffic Safety Education, Vol. XXIII, No. 1, October, 1975, pp. 15 and 17.

Haggard, E. A. Intraclass Correlation and Analysis of Variance. New York: Dryden Press, 1958.

Herbert, S. D. and Attridge, C. "A Guide for Developers and Users of Observational Systems and Manuals," American Educational Research Journal, 1975, 12, 1-20.

Hoty, C. "Test Reliability Estimated by the Analysis of Variance." Psychometrika, 1941, 6, 153-160.

Jones, Margaret Hubbard. Measuring the Outcomes of Driver Training: The USC On-Road Performance Test. Presented at the Transportation Research Board, January 25, 1977.

Kennedy, J. L. Driver Education and Training, Project, pp. 92. Report No. PH180-472, under contract FH-11-6561. Washington, D.C.: National Highway Safety Bureau, 1968.

Lybrand, William A., Carlson, Glenn H., Cleary, Patricia A., and Bower, Boyd H. A Study on Evaluation of Driver Education, pp. 210. Report National Highway Safety Bureau, 1968.

McGlade, Frances Stanley. An Evaluation of the Road Test Phase of the Driver Licensing Examination of the Various States: An Investigation of Current Road Tests and Testing Procedures, and the Development of a Valid and Reliable Road Test Based on Derived Implications. Dissertation, pp. 250. New York: New York University, 1960.

McKnight, J., and Hunt, A. G. Driver Education Task Analysis, Vol. I, Nov., 1970.

________. Driver Education Task Analysis, Volumes I, III and IV, 1970 and 1971.

Medley, D. M. and Mitzel, H. "Application of Analysis of Variance to the Estimated of the Reliability of Observations of Teachers' Classroom Behavior." Journal of Experimental Education, 1958, 27, 23-35.

________. "Measuring Classroom Behavior by Systematic Observation. In N. L. Gage (Ed.) Handbook of Research on Teaching. Chicago: Rand-McNally, 1963, pp. 247-328.

Michigan Department of Education. Michigan's Driver Education Evaluation Project. Lansing: The Department, 1978.

New York University. "Driver Education and Training--Plans for Evaluating the Effectiveness of Programs," pp. 95. Report No. PH 180-473, under contract FH-11-6560. Washington, D.C.: National Highway Safety Bureau, 1968.

Nolan, R. O., Vanosdall, F. E., and Smith, D. L., et. al. Driver Performance Research, Final Report, Vol. II, Guide for Training Observer/Raters in the Driver Performance Measurement Procedure. Prepared for National Highway Traffic Safety Administration, Contract FH-11-7627, Michigan State University, Department of Psychology, Highway Traffic Safety Center, February, 1973, p. vi.

Quenault, S. W. "Development of the Method of Systematic Observation of Driver Behavior," pp. 50. RRL Report LR 213. Crawthorne, Berks. (Gt. Brit.): Road Research Laboratory, 1968C.

Quensel, Warren P. "An In-Car Evaluation Instrument." Journal of Traffic Safety Education, January 1976, Vol. XXIII, No. 2, pp. 15-16.

________. "How to Measure Program Effectiveness," Journal of Traffic Safety Education, April 1976, Vol. XXIII, No. 3, pp. 6.

Rowley, Glenn L., American Educational Research Journal, Winter 1976, Vol. 13, No. 1, pp. 51-59.

Teal, Gilbert E., Truesdale, Sheridan L., and Fabrijio, Ralph A. Driver Education and Training, pp. 211. Report No. B2D68-575, Dunlap and Associates Inc., under contract FH-11-6559. Washington, D.C.: National Highway Safety Bureau, May 1968.

Uhlaner, J. E. Development of Criteria for Safe Motor-Vehicle Operation. Highway Research Board Bulletin 60, pp. 36-43. Washington, D.C.: Highway Research Board, 1952.

Vanosdall, F. E., et. al. Michigan Road Test Evaluation Study, Final Report, Vol. III. Prepared for National Highway Traffic Safety Administration, under contract MDL-75-002B, Michigan State University, Department of Psychology, Highway Traffic Safety Center, November, 1977.