utility programs for microcomputer operation and permits programming in Assembler, Fortran, or Basic. A library of subroutines is provided. With a good knowledge of RT-11 and either Fortran or Assembler, it is possible to write routines to acquire data from the transducers and store the data on diskettes. Using this approach, a program has been written to sample up to nine transducers every 100 ms and to store their outputs and a possible marker. This program has been used for transducer testing and early experiments.

Diagnostic software for checking the microcomputer was also purchased (DEC RXDP-11) in a convenient floppy-disk-based package, useful for exercising and checking the microcomputer, including memory, instructions, and interfaces. The diagnostic programs can be run in either stand-alone, chained, or interactive mode. Since the package is modular, other diagnostic programs can be added.

To run experiments in which many transducers are sampled and calculations are performed on-line, a dedicated system program is being developed. This system program will give the experimenter control over such parameters as sampling rate, channel to be selected, start-stop, and functions to be evaluated. The program will be modular so that it will be possible to add routines or modify existing routines without disturbing its overall framework. To save memory space it will not be necessary to load routines that are not required for an experiment. With this program, it is estimated that up to 15 channels can be sampled at 10-ms intervals.

CONCLUSION

This system has been operating on the road since the spring of 1977. Its success has been made possible by recent advances in microcomputer and data acquisition technology and continued progress in these fields will probably lead to higher sampling rates and relatively lower costs. In the meantime, it provides a unique opportunity to study behavior in actual driving situations in a quantitative manner.

ACKNOWLEDGMENT

We gratefully acknowledge the contributions of the following individuals and groups to this project: G. T. Chillcott, G. Stevens, H. Kuksin, M. C. Boyd, W. Fong, the DCIEM photoarts department, and shop.

REFERENCES

1. D. A. Attrood. Proposal for the Development of a Portable Data Acquisition and Analysis System for the Road Safety Unit of the Ministry of Transport. Dec. 1974.
2. H. C. Johnson and A. Presser. Automotive Doppler Radar Speed Sensor, RCA PE-585. RCA Labs, Princeton, NJ, 1973.
3. LSI-11, PDP-11/03 User's Manual. Digital Equipment Corporation, Maynard, MA, 1975.

Publication of this paper sponsored by Committee on Simulation and Measurement of Driving.

Notice: The Transportation Research Board does not endorse products or manufacturers. Trade and manufacturers' names appear in this paper because they are considered essential to its object.

Michigan's Driver Education Evaluation Project: Classroom Testing and In-Car Development

Kara L. Schmitt, Research, Evaluation, and Assessment Services,
Michigan Department of Education, Lansing

During 1977, the Michigan Department of Education evaluated statewide driver education programs. Rather than use the traditional approach of accidents and violations, the department investigated students' cognitive skills by means of objective-referenced tests. Previously, minimal performance objectives for both classroom and in-car portions of driver education had been developed. These objectives are considered basic for every Michigan program. After finalizing the objectives, appropriate test items were developed with the major focus on the classroom objectives. Following the pilot testing of the classroom items, 60 objectives were selected for statewide administration. Objective-referenced tests enabled not only the education department but also individual teachers to know the strengths and weaknesses of students and classroom environments. Each classroom objective was measured by five items; to attain an objective, a student would have to answer at least four items correctly. Statewide, the a priori criterion was that 80 percent of the students would attain each objective. Results based on approximately 100 000 students showed that only 13 objectives met the criterion level. These results indicate that there is a need for improvement in instruction of driver education. As skills became more advanced, attainment decreased. In the summer of 1977, an in-car measure was developed and raters were trained. The in-car test was administered to 30 students to determine its reliability. Forty students were used to validate the measure against the Michigan State University's Driver Performance Measure. Results for both studies were positive. Time, however, did not permit pilot testing this measure on a stratified sample of students.

Since the mid-1930s, Michigan has offered driver education to its youth. By 1954, approximately one-half of the state's school districts offered such instruction.

Then, in 1955, Michigan enacted a law mandating that all local school districts provide driver education. The law further states that all individuals under the age of 18 must have passed an approved driver education program before being eligible for a driver's license.

Although recognized as a leader in driver education, it was not until the last few years that any systematic research on this subject was initiated. In 1975, the Michigan Department of Education (MDE) received a 3-year grant from the Michigan Office of Highway Safety Planning under section 402 of the National Highway Safety Act to evaluate driver education. There were several events that prompted the MDE to write a proposal requesting such funds.

In 1972, the superintendent of public instruction was requested to report to the legislature on the effectiveness of current driver education programs (1). At that time, four questions were posed:

1. Do current driver education programs assure that the student acquires the knowledge and skills necessary to successfully pass 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?

Using questionnaires and interviews, the 1972 study addressed only the fourth question. However, the results of the study were tentative and provided only possible relationships between driver education and road safety. Therefore, additional research was necessary.

In part, the present project is an attempt to answer the general concepts of the first two questions. The third question is difficult to respond to since there is no control, or comparison, group presently available. Although the 1972 report did recommend that a control population be formed through legislation, this has not occurred. Present records indicate that almost all 18-year-olds have received driver education training from an approved program.

A second factor, which prompted this study, was the MDE's consideration in 1974 of asking the legislature to increase the driver education reimbursement. Public schools currently receive $30.00 for each student who completes a minimum of 24 hours of classroom and 4 hours of road instruction—the minimum for issuing a certificate is still the successful completion of at least 30 hours of classroom and 6 hours of road instruction. However, at that time, it was decided that more information was required before any recommendation be made regarding this issue.

Finally, it was the impression of the MDE staff that schools and individual teachers did not have any objective method by which to determine program effectiveness. Intuition and "gut level" feelings were being applied in the decision of what should be taught and how. This impression is emphasized by Chapman (2, p. 95), who states that "for too long, an unnecessary burden has been placed on those responsible for driver preparation, forcing them to conduct programs in terms of performance objectives they have to derive intuitively."

The department decided that the use of accident records, points, injuries, and so forth was not necessarily the most appropriate procedure to follow. Previous research using this method has generally yielded conflicting or inconclusive results. Brody maintains that ". . . if we seek to evaluate driver education in terms of accident reduction, we are confronted with so many variables, known, unknown and highly variable if not unpredictable, that we become enmeshed in an endless chain of proof" (3, p. 85). If indeed younger drivers do have a greater percentage of accidents per kilometer driven, it may be no more than a cultural artifact. It may be related to inexperience and early learning errors rather than to age or emotional maturity.

The ultimate goal of driver education is to produce collision-free drivers; however, the total elimination of crashes and violations by young drivers is an unrealistic goal. The way to assess the achievement of this goal may not be through drivers' records because, as R. Zylman states, "What appears on the driver's records may have nothing to do with driver behavior" (4, p. 7). He continues, "It is clear that driver's records as they appear in the DMV [Department of Motor Vehicles] files are not a valid measure of driver behavior. The problem has its origin in the fact that data are being used for purposes for which they were not intended."

For these reasons, the MDE decided to evaluate its driver education programs with objective-referenced measures of both students and classroom situations. Objective-referenced tests are designed to differentiate between students who have and have not mastered stated objectives. The measures are not aimed at ranking individuals or comparing them with each other as is true in norm-referenced testing. Objective-referenced tests focus on a student's achievement relative to a predetermined criterion or standard of quality, while normed tests focus on a student's standing relative to a given group. Besides using these measures to ascertain which students have mastered the material, it is also possible to use the results in conjunction with an analysis of the classroom delivery system. Theoretically, following instruction, all students should master all objectives taught. If the class as a whole does not perform well, a teacher could look at the way in which the material has been presented and make the necessary changes to improve the achievement level.

PROJECT DESIGN

The basic design for the project involved five stages:

1. Develop, review, and revise performance objectives for both the classroom and in-car experiences;
2. Develop test measures to match each of the performance objectives;
3. Pilot test, review, and revise the classroom items—selecting those appropriate for a final battery—and determine the reliability and validity of an in-car measure;
4. Administer the final classroom measure to all students; and
5. Analyze the data.

Performance Objectives

The initial step in the evaluation project was the development of basic objectives that should be included in all driver education programs regardless of program format or type of school (public, private, parochial, or commercial) offering the course. The first draft of the objectives was written by one of the driver education consultants on the MDE staff.

Because of the potentially wide variability across the state, it was decided that as many driver education specialists (teachers, curriculum specialists, professors) as possible should be included in the review and revision

of the objectives. It was assumed that their pooled opinions would represent the general principles included in all Michigan driver education programs.

During the first year of the project, the objectives were reviewed by over 250 individuals. The following four criteria were used in this review:

1. They should be minimal or basic to all programs. Knowledge of the history of driver education would not be considered a basic skill.
2. They must be specific rather than global or general. An objective such as "A student will be able to drive safely at all times" is too general to be of any value. A better objective is "A student will know the meaning of roadway signs by their shapes."
3. The wording of the objectives has to be clear to those using them. Incorporating terminology that is vague will only cause teachers to interpret the meaning in a variety of ways; thus, consistency is lost.
4. The objectives must be measurable. If an objective cannot be tested by some procedure, then it serves no useful purpose.

These persons also rated the objectives according to their importance (priority) in producing safe drivers (e.g., how to plan a trip was not considered very important) and to the similiarity with which they are taught statewide. For example, freeway driving receives greater attention in the metropolitan areas than in the rural areas, but the procedures for making a right turn should be emphasized in all geographical areas.

The final set of objectives included both classroom (cognitive) and in-car (psychomotor) skills. In the first category, the students would be tested by paper and pencil tests; in the second, students would drive a vehicle and would perform certain specified maneuvers. A trained observer would rate the students' performance on each of the selected objectives.

It was decided that there would not be any objectives dealing with the affective domain. The reason for this is that most measures of this nature are very superficial, and students are capable of responding in the manner deemed appropriate for the circumstances. Thus, it is difficult to obtain any accurate measure of a student's true attitudes or beliefs.

Classroom Performance Measures

With the finalization and general acceptance of the performance objectives, the item-writing process began. Four workshops in which 120 individuals participated were conducted across the state. At each workshop, several hours of instruction were provided on how to write proper objective-referenced test items. Some of the important factors that the item writers were taught included:

1. The language should be at the sixth-grade level except for specific driver education terminology. If uncertain about a word, writers should try to use a simpler one.
2. Items should be presented in a straightforward manner. The intent is not to trick students.
3. The items should measure a single concept and should be neither too long nor too involved.
4. When reasonable, four responses (foils) per item should be included. The responses should be appropriate for the question being asked.

The principal criterion for good items is that each item directly measure the objective for which it is being written. As Popham and Husek (5, p. 133) point out, the objective-referenced test writer's "chief rule is to make sure the item is an accurate reflection of the criterion behavior. Difficult or easy, discriminating or indiscriminate, the important thing is to make the item represent the class of behaviors delineated by the [objective]."

The goal was to have between 10 and 15 items per objective; the results were 0 to 115 items. There was one objective which, unfortunately, was written in vague terms and could not be clearly interpreted by the item writers. The objective covering sign shapes was so easy that 115 items were constructed.

Following the workshops, the items were reviewed by the MDE staff and then submitted to a technical support contractor. The goal of the technical support contract was to obtain at least 10 good items for each of the 152 classroom objectives by either revising items written during the workshops or constructing new items using the teacher items as prototypes. In most instances this goal was met; however, there were a few objectives so limited in scope that the development of 10 items was not feasible. The objective for which no items were written was dropped from consideration because of the difficulty in interpreting its intent.

The MDE reviewed each of the items prepared by the contractor and then had a meeting with the contractor to determine the manner in which the items would be pilot tested. Following this review, four additional objectives were dropped because of problems in writing items which would not be biased or too obvious. Therefore, a total of 147 objectives tested by 1366 items were prepared for pilot testing.

The items were packaged in 20 test booklets of approximately the same length. Each test book covered about eight objectives, each of which was measured by approximately 10 items.

Pilot Testing

Schools participating in the pilot testing of the classroom test items numbered 126. The schools were selected randomly from the four most common program formats: two-phase, three-phase range, three-phase simulator, and four-phase. The number of schools selected in each category was based upon the total percentage of schools in the state offering the various formats. Programs ending between July 26 and August 27, 1976, were eligible for selection. Consideration in the selection process was also given to the size of the city (metropolitan, urban, rural) as well as geographical location.

In addition to the summer posttesting tryouts, eight schools participated in pretesting some of the high priority items. These schools administered the tests during the first few days of their fall programs. The information obtained from the pretests enabled the MDE to be aware of students' performance prior to any formal classroom training.

Teachers in both the pretesting and posttesting samples were requested to provide the MDE with comments concerning both the general aspects of the tests as well as comments about specific test items. These comments would be included during future review sessions.

Selection of Items for Statewide Testing

The tests were scored by Michigan State University, and results were sent to each of the participating teachers. The university also performed the necessary test analyses. Since these were objective-referenced tests, the analyses were computed on each of the objectives and

corresponding test items. The analyses computed included a foil analysis, a KR_{20}, Livingston internal consistency (6), point biserials, biserials, and item discrimination.

In mid-October 1976, a review meeting was held in which 50 driver education teachers and specialists reviewed the test results, statistics, and teacher comments. Suggestions were made regarding revisions and deletions of items. Also, input was provided on which objectives should be included in the final battery.

Another review session was held with the contractor and MDE staff. During this time, decisions were made concerning the final test and format. Decision rules for dropping or revising items depended on the content of the objectives and teachers' review. Items for objectives which were rated as commonly taught required a higher P-value (percent of students answering correctly) than those less commonly taught. Foils which did not operate well were revised where possible. Items were dropped in cases where the point biserials were too low. A special effort was made to simplify the vocabulary and reading level of all items retained. Several entire objectives were dropped because of measurement difficulties, high pretest results, or the inadequacy of the objectives themselves.

The final tests covered 60 basic objectives generally considered by driver education teachers as essential for effective programs. The five best items for each of these objectives were chosen for inclusion in the test packages.

The choice of five-item tests was based upon several considerations. The first was the minimum number of items judged necessary to provide a reasonably adequate measure of a student's attainment of an objective. A second was the amount of testing time available—the average student should be able to complete the test in one classroom period. However, since the test was untimed, the slower reader would not be penalized.

Of the five items per objective, a student must answer correctly four or five items to be considered as having mastered the objective. Since the classroom portion of the evaluation was in a multiple-choice format, it was necessary to minimize the possibility of misclassification. Therefore, a passing score was selected so that an objective would be attained by guessing no more than 2 percent of the time. According to Millman (7), the passing score associated with no more than a 2 percent chance of its being obtained by random guessing is 80 percent, or four out of five in this situation.

Four test booklets were designed. Each one covered 15 objectives measured by five items for a total of 75 items. The 15 objectives per booklet were randomly selected from the total set of 60 objectives. This was done so that the booklets would be comparable in the areas being tested.

Following the packaging of the booklets, a readability analysis was computed on each of the items as well as on the four test booklets. The goal was to have the average reading level of each book at the sixth grade. The four booklets had an average reading level of 6.3, 6.2, 6.4, and 6.3. Therefore, the reading level was consistent among the four forms and was not likely to have an adverse effect when students' performance was analyzed.

In-Car Performance Measure

During the summer of 1977, the MDE, with assistance from a Michigan State University doctoral student in traffic safety, began designing an in-car road test. The intent of the instrument was to measure certain in-car performance objectives; however, those dealing with parking and emergency situations were to be omitted from the final measure.

The initial activity was the design of an appropriate route. The route would need to include a variety of driving situations—city, expressway, and residential—yet not require too much driving time. Certain other driving elements had to be included, such as turns from and to one-way and two-way streets. A further consideration was that student drivers would not be placed in an intentionally dangerous situation; that is, they would not be requested to make unsafe maneuvers or be forced off the road and try to regain control. The eventual route encompassed all desired objectives. It covered approximately 17.6 km (11 miles) and required about one-half hour of driving time.

The route was divided into areas of specific and general observations. The specific areas were designated as Location(s) of Performance Evaluation (LOPEs). While in these areas, the raters would pay close attention to definite driver behaviors. No recording of observations would occur at these points; instead, the recording plus general observation would occur at other times. Also, there would be a warmup period of several blocks during which both the rater or raters and the driver could become acquainted, thus reducing any apprehension or tension felt by the driver.

Specific directions, as well as points where directions were to be given, were written. Strict adherence to the directions and locations for giving them was necessary for appropriate driver reaction. Directions would be given with sufficient time for the driver to make all requested maneuvers safely.

The actual in-car test was divided into two main sections: vehicle familiarization and driving. The vehicle familiarization section was a checklist for indicating whether or not an individual knew various parts of the vehicle, e.g., information gauges, starting and control devices, and safety devices. Also included were observations of a driver's preignition control tasks: starting the engine, putting the vehicle in motion, stopping the vehicle, and securing it.

The driving section consisted of a series of observations to be made by the evaluators. There are three main concepts involved in all driving—search, speed, and direction. Included in each of these points is the element of timing. Although timing is an important element of driving, it was included in the evaluation only when an objective specifically mentioned it. Finally, there were certain areas in which the drivers were judged on whether they yielded to pedestrians, or vehicles, or both. For each of the Specific Performance Objective Test Sites (SPOTS), an indication of satisfactory and unsatisfactory behavior was listed. These descriptions would assist the evaluators while making their judgments as to the driver's performance. The last page of the driving unit included another type of checklist related to how well the driver obeyed traffic signs, signals, and lane markings.

Five individuals were selected to be raters for the driving measure. Each individual had completed the training necessary for riding with unlicensed drivers. These five persons, along with an MDE member, participated in a 3-day training session. The route, directions, and meaning of terms were explained during this training. One of the most crucial aspects of the training was the familiarization of the route. The key to a successful evaluation was knowing the route and what to observe. Frequent reviews of the route, via actual driving, were included during the training. Also, some practice trials with student drivers were made so that raters' perceptions and scoring could be reviewed and discussed.

Following the training session, the agreement be-

tween ratings on practice drives was reviewed to determine if any of the raters were less qualified than others. The various pairings of raters indicated that agreement was quite high; therefore, the actual pairings for the test situation would not influence results.

Reliability of In-Car Measure

During the first two weeks of August 1977, the reliability study was carried out. An attempt was made to avoid Mondays and Fridays because of the atypical driving situation on these two days. Students from seven local school districts were selected randomly for participation. The school districts had recently completed a driver education program. The students selected had passed the course, but were not yet licensed drivers. Letters were sent to the parents or guardians of 80 students requesting permission for them to participate. The study required at least 45 students. The additional students were included because of possible time conflicts, vacations, or failure to return the permission forms.

The chosen design was such that each of 30 students would be evaluated by four raters. The extra 15 students would be used as alternates in case some of the 30 students were unable to complete the test.

Of the six trained raters, four were chosen to test the 30 subjects. Because there were no significant differences in the ratings for the practice trials, the four were selected randomly. The other two raters observed the alternates.

Validation of In-Car Measure

Following the analyses for the reliability study, plans for a concurrent validation were initiated. Most driving tests rely upon content validation rather than including the additional step of concurrent validation. The department was fortunate enough to have a criterion measure that could be utilized for establishing this additional validity.

The Highway Traffic Safety Center (HTSC) at Michigan State University (MSU) had developed a carefully designed and validated road measure (criterion instrument) for use in the concurrent validation of in-car road tests. The HTSC procedure provided a criterion of safe and skillful driving performance necessary to validate the department's road test. The Driver Performance Measure (DPM) is considered as a research tool to "serve as a basis for and as a means of validating practical, simpler testing procedures which could be used by teachers and examiners" (8, p. 26).

The validation study took place during the months of October and November 1977. A total of 40 students, randomly selected from students who had successfully completed the Lansing summer program, were assigned a particular date and time. As in the reliability study, alternates were chosen in case some students were unavailable at the requested time.

Four individuals trained in using the DPM were included in this study; three of the MDE's raters were used. Both the MDE and DPM raters received additional training so that they were familiar with the route, the measures, and the characteristics to observe. Several practice trials were provided.

Each subject drove the department's route once and the DPM route twice. One-half of the students drove the MDE route first; the other half drove the DPM route first. This procedure was utilized to alleviate any possible practice effects. Also, during the DPM, there were two raters per drive. After the first drive, the raters switched positions, from front to back seat and vice versa. This would reduce any effect either position might have on the observations.

Internal consistency measures were computed on both the DPM and MDE tests. Both measures included scores on students' speed, search, and directional control. In addition, the DPM had an overall rating (PATTERN) for each point of observation. This score was determined by whether or not the driver performed safe or unsafe maneuvers in relation to the environment. The PATTERN score was not related directly to the other three scores. Finally, a correlation between driving scores on the MDE and the DPM was computed. Both total score and element scores were correlated. Since there were four DPM ratings per subject, these correlations were made using the average DPM score.

Testing

Beginning January 17, 1977, all students enrolled in a driver education program at a public, private, parochial, or commercial school were required to take one form of the classroom test at the conclusion of their course. About 3 weeks prior to the end of each program, teachers received the required number of tests and answer sheets as well as instructions for administration. By December 15, 1977, approximately 140 000 students had been tested.

In addition to those students who completed the test at the conclusion of their course, some 4000 students took the test prior to any formal instruction. Teachers were requested to administer the pretests (the same tests used in the posttesting) to the students during the first or second day of class. The pretest information would provide the MDE with an idea of students' cognitive knowledge prior to receiving their formal driver education training.

Because the driver education evaluation was not intended to penalize students with reading difficulties, teachers were informed that the test was not to be timed. Therefore, students could have as much time as needed to complete the test. Also, the teachers could answer any questions regarding the meaning of words, if the assistance did not give away the correct answer. In many instances, the tests were read to the students; some schools even provided cassette recordings of the test or audiovisual materials.

Teachers were requested to submit comments to the MDE after they had administered the evaluation tests. The department received over 400 general comments, which were generally favorable toward the test construction and intent of the project.

Less than a month after a class completed the test, the teacher received two reports. The first indicated how well each student did on each objective, and the second showed the percentage of students in the classroom answering 0 to 5 items correctly for each objective. By using the objective-referenced testing procedure, it was hoped that teachers would use the information provided to revise their curriculum where they deemed it necessary.

STATEWIDE ATTAINMENT

Throughout the year, the MDE received periodic reports on the statewide attainment of the 60 basic objectives. The final report was based on approximately 25 000 students per test form. The percentage of students correctly answering four or five items for each objective is shown in Table 1. The criterion level established was that 80 percent of the students would correctly answer at least four items. Table 1 shows that only 13 of the 60 objectives were attained—less than 22 percent of

the basic objectives. Even if attainment were lowered to 75 percent, only an additional nine objectives would be met. The number and percent of objectives and cumulative objectives achieved by students are shown in the following table:

Students (%)	Number of Objectives Achieved	Objectives Achieved (%)	Cumulative Percentages
80-100	13	22	22
75-79	9	15	37
50-74	32	53	90
25-49	5	8	98
0-24	1	2	100

For students to have achieved an objective, they must have answered correctly at least four of the five items.

The classroom objectives are divided into six units. The first two represent the very basic tasks of vehicle familiarization and basic control tasks. Of the 43 objectives tested from these two sections, only nine reached the desired level of attainment and eight of these were in vehicle familiarization. This unit covers areas of the driving compartment, starting and stopping the vehicle, driving fundamentals, and traffic control. The objective with the poorest achievement (49 percent) in the first unit dealt with signs and their respective colors. Putting the vehicle in motion had the highest

Table 1. Attainment rates for 60 objectives used in the Michigan driver education project.

Objective	Description	Attainment: Pretest (%)	Attainment: Posttest (%)	Difference (%)
Vehicle familiarization				
A2.2	Preignition control procedures	58	70	12
A2.3	Starting engine	55	67	12
A2.4	Putting vehicle in motion	80	89[a]	9
A2.5	Stopping vehicle	69	73	4
A2.6	Securing vehicle	75	82[a]	7
A3.2A	Steps for right turn	38	60	22
A3.2A	Steps for left turn	33	41	8
A3.3	Vehicle positioning for turning left or right	50	67	17
A3.4A	Natural forces when rounding a curve	41	61	20
A3.4B	Compensation for effects of natural forces	53	70	17
A3.5	Effects of gravity going up or down hill	55	62	7
A3.6	Maintaining proper speed control	74	81[a]	7
A4.1A	Traffic sign shapes with their purpose	71	88[a]	17
A4.1B	Traffic sign shapes with driver action	55	79[b]	24
A4.1C	Traffic sign shapes with colors	25	49	24
A4.2A	Traffic signals with their purpose	70	80[a]	10
A4.2B	Traffic signals with driver action	68	79[b]	11
A4.3A	Pavement markings with their purposes	76	86[a]	10
A4.3B	Pavement markings with driver action	68	80[a]	12
A4.4	Driver action for traffic control persons	71	83[a]	12
Basic control tasks				
B1.1	Vehicle movement at intersections	54	63	9
B1.2A	Potential conflict at intersections	59	72	13
B1.2B	Reducing risks at intersections	64	77[b]	13
B1.3A	Entering freeway	30	57	13
B1.3B	Exiting freeway	43	61	18
B1.4A	Path of travel on freeway interchanges	72	78[b]	6
B1.4B	Potential conflicts at freeway interchanges	60	67	7
B2.1	Maintaining a space cushion	71	85[a]	14
B2.2	Speed and directional control on space cushion	54	75[b]	21
B2.3	Establishing proper following distances	51	67	16
B2.4	Vehicle blind spots for various situations	47	70	23
B2.6	Kinetic energy on stopping distance	8	22	14
B4.1	Conditions requiring lane changes	50	69	19
B4.2	Procedure for lane changes	44	73	29
B5.1	Body and hand positions for backing	42	70	28
B5.2	Vehicle handling when backing	36	51	15
B6.2	Overtaking and passing another vehicle	44	72	28
B6.3A	Legal overtaking and passing on the left	70	79[b]	9
B6.3,4	Illegal overtaking and passing on right/left	51	60	9
B6.4	Legal overtaking and passing on right	29	51	22
B6.5	Overtaking and passing a school bus	41	55	14
B7.2	Minimizing hazards when being passed	40	48	8
B8.5	Communicating with other vehicles	70	75[b]	5
Driver fitness tasks				
C1.2	Compensation for visual impairments	59	77[b]	18
C3.1,2	Minimizing distractions (inside and outside)	78	84[a]	6
C4.2	Compensation when under emotional situations	61	76[b]	15
C6.1	Physical effects of alcohol	74	82[a]	8
C6.2	Psychological effects of alcohol	42	57	15
Intermediate control tasks				
D1.2B	Increased speed on friction and impact	27	51	24
D1.6B	Hazardous features of freeway engineering	48	61	13
D2.1	Parallel parking on incline	21	40	19
D5.6	Social responsibilities of drivers	70	82[a]	12
Advanced control tasks				
E1.1	Hazardous situations for night driving	35	73	38
E1.2	Weather conditions and hazardous driving	74	84[a]	10
E1.2	Precautions to take for adverse weather	38	70	32
E1.5	Factors reducing availability of friction	35	59	24
E2.1	Driver response for emergency situations	30	61	31
E3.1A	Vehicle warning lights (malfunctions)	59	65	6
E3.3	Steps to take with vehicle malfunctions	56	66	10
Vehicle				
F2.4	Procedures to follow in traffic accidents	35	45	10

[a] 80 percent of the students answered 4 or 5 items correctly.
[b] 75-79 percent of the students answered 4 or 5 items correctly.

attainment (89 percent), both in this unit and in the entire test.

The second unit, basic control tasks, includes the concepts of being passed and passing, backing, lane changes, being followed and following, and intersections. The only objective in this unit that was attained dealt with maintaining a space cushion. The objective on the effects of kinetic energy on stopping distance had an attainment of only 22 percent. The items measuring this objective involved the notion of what happens when the speed of a vehicle is doubled. Most students responded that the stopping distance doubled as the speed doubled.

The best unit with regard to high achievement was on driver fitness tasks. This unit includes the topics of drugs, alcohol, risk taking, attitudes, distractions, and physical attributes. Attainment ranged from 57 percent (psychological effects of alcohol) to 84 percent (minimizing distractions—inside and outside). Students did better with the physical rather than psychological effects of alcohol.

There were four objectives included from the intermediate control tasks unit. One objective, social responsibilities of drivers, had an achievement level of 82 percent. The other objectives were below 65 percent attainment.

The unit on advanced control tasks includes topics of what to do in an emergency or in adverse conditions. Seven objectives were tested, but only one had an attainment of over 80 percent. The other attainment levels ranged from 59 to 73 percent.

The final unit on the vehicle deals with car care, purchasing a vehicle, insurance, and accidents. Only one objective was considered basic enough to test. This one dealt with the procedures to follow when involved in a traffic accident. Only 45 percent of the students were able to answer at least four of the five questions.

From the attainment rates, it appears that most of the course work covered the sections on vehicle familiarization and driver fitness tasks. As the skills became more advanced, attainment decreased. Thus, it appears that more time is required in the classroom or that attention should be focused on the more critical skills.

In November 1977, a meeting of 35 driver education specialists was held to review the items and corresponding objectives. Their task was to determine the probable reason for poor attainment. The main question to be answered was whether lack of attainment was due to poor items or limited instruction.

The reviewers' overall impression of the objectives tested was that they were appropriate, measurable, and teachable. However, because of the shortness of most programs, many objectives were deleted or only mentioned briefly. In some areas, such as natural forces (friction, gravity), the teachers themselves may not have fully understood the concepts.

Since attainment was chosen to be at least four correct answers per objective, even with one poor item, students should have been able to answer the other four correctly if they knew the material. There were a few instances in which two or three items were inappropriate or incorrect; therefore, attainment would be expected to be low. Of the 60 objectives tested, 28 had attainments below 70 percent. Of these 28, 15 were judged to have low attainment because of limited or insufficient teaching. A few of the general areas, which did not appear to be emphasized in the courses, were emergency procedures, accident involvement, natural forces, parallel parking, and freeway driving.

Although the tests were not to be timed, some teachers may have stopped the students after a given amount of time. If this occurred frequently, a decrease in attainment should have been found near the end of each booklet. This did not seem to be the case. An inspection of the results for each booklet revealed that there was no consistent pattern of decreasing attainment levels in the latter sections of the booklets. The most striking example of a nondecreasing trend occurred in booklet three. Some 82 percent of the students attained the last objective, social responsibilities of the driver; while only 22 percent attained the next to the last objective, kinetic effects on stopping.

The pretest attainment levels along with the difference between pretest and posttest results are also shown in Table 1. The results showed an improvement in all the objectives; therefore, some acquisition of knowledge occurred during the driver education programs. The pretest attainments ranged from 8 to 80 percent and the posttest results ranged from 22 to 89 percent with the same objectives being the highest and lowest in both instances. Fifteen objectives had pretest attainment of 70 percent or better. There were 36 objectives (60 percent) in which at least one-half of the students answered four or five items correctly in the pretest.

The difference between pretest and posttest results ranged from 4 to 38 percent with an average increase of slightly more than 15 percent. The objective showing the greatest improvement (from 35 to 73 percent) covered hazardous situations for night driving. The smallest gain (from 69 to 73 percent) was on how to stop a vehicle.

For those objectives that reached the posttest criterion of 80 percent, the increase from pretest to posttest was about 10 percent. This would tend to indicate that on objectives for which desired posttest attainment was achieved, students knew the material fairly well prior to beginning the course.

DRIVING PERFORMANCE MEASURE RESULTS

The other performance measure developed was for the observation of actual driving behavior. During the first few weeks of August 1977, 30 students were tested for the purpose of determining the reliability of an in-car measure and the raters. Four raters evaluated each of the students. Rather than combining the results of the four individuals, a separate reliability factor was obtained for each of the test components for each rater. The following table shows the test reliabilities:

Test Component	Rater 1	Rater 2	Rater 3	Rater 4
Sum	0.92	0.90	0.86	0.87
Drive	0.91	0.88	0.86	0.87
Search	0.77	0.67	0.83	0.84
Speed	0.80	0.75	0.76	0.61
Direction	0.84	0.87	0.78	0.81
Familiarization	0.67	0.71	0.59	0.34
Signs	0.42	0.76	0.37	0.44

The sum represents the entire test and includes 147 observations. The drive component consists of the actual driving skills and includes search, speed, and directional control. Familiarization tasks involve whether students know the location of various parts of the vehicle as well as how to start and stop a vehicle. The final section, signs, includes students' attention to and compliance with the designated signs and pavement markings.

All but five of the reliabilities obtained were greater than 0.60, which is quite acceptable regardless of the type of research and is very good for a driving measure. The values derived in a study by Jones (9) and in the

Table 2. Pearson correlation of the MDE in-car measure.

Test Component	Sum	Drive	Search	Speed	Direction	Familiarization	Signs
Sum	1.00						
Drive	0.98	1.00					
Search	0.86	0.86	1.00				
Speed	0.83	0.84	0.67	1.00			
Direction	0.77	0.81	0.47	0.52	1.00		
Familiarization	0.50	0.34	0.40	0.45	0.04	1.00	
Signs	0.43	0.36	0.26	0.10	0.49	-0.07	1.00

Table 3. Validity of the MDE driving component.

	Driver Performance Measure				
MDE	Sum	Search	Speed	Direction	Pattern
Drive	0.46[a] (0.50)	0.35[a]	0.35[a]	0.32[b]	0.44[a] (0.51)
Search	0.51[a]	0.56[a] (0.60)	0.23	0.29[b]	0.42[a]
Speed	0.32[b]	0.21	0.32[b] (0.40)	0.16	0.36[a]
Direction	0.31[b]	0.10	0.34[b]	0.32[b] (0.36)	0.33[b]

[a]$p < 0.01$. [b]$p < 0.01$.

Michigan Road Test Evaluation (November 1977) by Vanosdall and others—the latter unpublished—were equal to or less than those obtained in the present study. The lower values for familiarization were due to a lack of variance within the items—71 percent of the items had means of 0.9 or better and 44 percent had means of 1.0; the items were scored as either 0 or 1. The small number (9) of items for the signs section also affected the reliability results.

In addition to the reliabilities of the measure, rater agreement was calculated for each of the test components. Since the raters were not always paired with the same person, these correlations were computed for pair one (disregarding individuals) on drive one and pair two on drive two. The results are shown below:

Test Component	Pair 1	Pair 2
Sum	0.86	0.83
Drive	0.80	0.80
Search	0.67	0.79
Speed	0.72	0.51
Direction	0.75	0.50
Familiarization	0.83	0.52
Signs	0.49	0.54

The correlations obtained showed a high degree, in most instances, of agreement between the pairs of raters. In all instances the correlations were significant at $p < 0.01$. In other words, both raters saw the same behavior and rated it in similar fashion. These results verify that the training program was successful.

Because acceptable reliabilities were obtained, it was possible to proceed with a concurrent validation study. This study was conducted in October and November 1977, after students had been driving for at least two months. Most students (85 percent) had already obtained their driver license. Each of the 40 students was rated by one MDE rater and two DPM raters. In addition, the DPM route was driven twice so that there were actually four observations provided on the DPM.

Interitem reliabilities were again computed for the MDE test and found to be consistent with those of the reliability study. The correlations ranged from 0.54 for familiarization to 0.89 for drive. Except for familiarization and signs, all correlations were above 0.70.

The intercorrelations of the total test score and each of the six subscales are presented in Table 2. Scores on the drive component, which comprises approximately 60 percent of the total test, correlated almost perfectly (0.98) with the total test score. Each of the elements that were a part of the drive correlated above 0.80 with drive. However, both the familiarization and signs elements had low correlations with the total test score (0.50, 0.43) and even lower correlations with drive (0.34, 0.36). This would indicate that these two components were not necessarily measuring the ability to drive properly as measured by the MDE drive. For example, students may know where each of the parts of a vehicle are, but may or may not know how to operate a vehicle correctly. Finally, the correlation between sign and familiarization was -0.07, indicating no relationship between the two components.

The correlations between scores obtained on the DPM and the MDE were computed to ascertain the validity of the MDE measure. The mean score of the DPM was used in the correlation since there were four observations made per student. The results are shown in Table 3.

According to the authors of the DPM, the pattern score is the most important in determining whether or not an individual can drive safely. This element correlated 0.44 with the MDE drive component, which was significant at $p < 0.01$. In fact, the MDE drive components were significant at $p < 0.05$ when correlated with the corresponding DPM components. Of the 20 correlations obtained, only four were found to be nonsignificant. The correlations obtained in this research are comparable to those obtained when the DPM was used to validate a new Michigan state road test by Vanosdall and others.

If only one DPM rater had been used, the correlations between the MDE and DPM would have been equal to the values shown in parentheses in Table 3. In all instances, the MDE elements would have been significant at $p < 0.01$ when correlated with the appropriate DPM elements.

DISCUSSION

The goal of the driver education evaluation project was to answer several questions regarding Michigan's driver education programs. The major question was: Are current driver education programs achieving the classroom performance objectives? The answer was obtained by the use of objective-referenced tests. After 1 year of testing, the MDE now has a clearer idea of what is being taught in driver education programs and thus can provide suggestions or assistance in improving the attainment of the classroom performance objectives.

The development of appropriate performance objectives should enable the state to obtain consistency across

all programs. The objectives will, at least, provide teachers with more complete knowledge of what should be taught. In turn, the consistency in performance objectives should be of value for future evaluation studies of driver education programs.

The statewide results indicate that the material considered basic for all programs is not being attained as expected. Using the a priori criterion of 80 percent of the students answering correctly at least four items per objective, only 22 percent of the objectives were met. If the criterion were lowered to 75 percent, then 37 percent of the objectives would be met; reduction to 70 percent attainment would mean that 53 percent of the objectives were met. Further reduction in the criterion level would be inappropriate given the way the tests were constructed.

In fact, when one remembers the basic premise of the test construction, the theoretical goal would be that everyone should be able to answer correctly every question. The objectives were selected by a representative sample of driver education teachers and specialists as being basic; the questions were constructed by driver education experts as being elementary and not tricky; and the average reading level of each test booklet was sixth grade. These factors were incorporated to assure that a criterion of 80 percent would not be unrealistic.

One positive aspect of the research was that there was always an improvement from the pretest to the posttest scores. Some learning is obviously occurring in driver education and this would indicate that driver education should not be eliminated. However, on those objectives that reached the posttest criterion level, there was less improvement from the pretest to the posttest scores than for the other objectives. This would indicate that, for these particular areas, students were already familiar with the topic prior to entering the program. It might be advisable for teachers to administer a pretest to their classes and advance students from their entrance level. This might mean more time could be spent on the advanced skills.

Another factor to consider is the time generally available for teaching—less than 30 hours in typical programs when all nonteaching duties are considered. The 30 hours of classwork may be insufficient for the instruction considered necessary. Even in longer programs, teachers may be merely reiterating the same concepts as in short courses rather than working on the advanced skills. More careful attention should be paid to this variable along with an in-depth review of what is being taught in the longer programs.

Comments from the posttest review pointed out that the most frequent reason for low attainment on the tests was probably the limited or absence of teaching about many areas. This could be due to the time constraints imposed by the 30-hour classroom minimum requirement or because teachers were unfamiliar with some of the material. There were only a few instances in which the test items were viewed as the reason for failure.

It is somewhat paradoxical that both the students and teachers considered the test to be fairly simple. Some even suggested that the tests should be made more difficult or that the correct answers be less obvious. However, these comments did not coincide with the results obtained.

The driving measure was shown to be both reliable and valid. However, because of the time constraints imposed upon the project, it was not possible to select a stratified random sample of students for actual test administration of the in-car measure.

The results of the classroom testing should not be construed to mean that driver education programs are a waste of time; but rather that there is a need to take a closer look at what students are being taught and how the 30 hours of classroom time are being used, as well as the preparation that teachers have received. Perhaps it is necessary for the MDE and the local school districts to work together to provide the necessary preservice and inservice training so that teachers will be better able to instruct their students. Also, a closer examination of how best to present the material in the given time frame would be important. Finally, a more extensive examination of the longer or full-semester programs should be implemented.

Since the department and the local school districts now have some concrete data on the achievement of students' cognitive knowledge, some changes may be possible. Although the results obtained provide a clearer picture of Michigan's driver education program, there are still questions that need to be answered. However, at this time, the information contained herein is the best information available regarding classroom performance.

ACKNOWLEDGMENT

The opinions, findings, and conclusions expressed in this publication are and not necessarily those of the Michigan Office of Highway Safety Planning or the U.S. Department of Transportation. This paper was prepared in cooperation with the Michigan Office of Highway Safety Planning and the U.S. Department of Transportation, National Highway Traffic Safety Administration.

REFERENCES

1. Report on the Effectiveness of Current Driver Education Programs. Michigan Department of Education, Sept. 1972.
2. R. Chapman. Plan for Critical Appraisal of Driver Education Programs. In Proc., National Driver Education and Training Symposia (Public and Nonpublic Schools, Dec. 1-4, 1968; Commercial Driving Schools, Jan. 26-29, 1969), El Segundo, CA, Institute for Educational Development, 1969, 275 pp.
3. L. Brody. Plan for Critical Appraisal of Driver Education Programs. In Proc., National Driver Education and Training Symposia (Public and Nonpublic Schools, Dec. 1-4, 1968; Commercial Driving Schools, Jan. 26-29, 1969), El Segundo, CA, Institute for Educational Development, 1969, 275 pp.
4. R. Zylman. Evaluating Driver and Traffic Safety Education Programs: It May Be Tougher Than You Think. Journal of Traffic Safety Education, Oct. 1973, pp. 7-8.
5. W. J. Popham and T. R. Husek. Implications of Criterion-Referenced Measurement. In Educational and Psychological Measurement (D. A. Payne and R. F. Morris, eds.), General Learning Press, Morristown, NJ, 1975, pp. 129-139.
6. S. A. Livingston. Criterion-Referenced Applications of Classical Test Theory. Journal of Educational Measurement, Vol. 9, 1972, pp. 13-26.
7. J. Millman. Passing Scores and Test Length for Domain-Referenced Measures. Review of Educational Research, Vol. 43, 1973-1974, pp. 205-216.
8. T. W. Forbes and others. Driver Performance Measurement Research. Journal of Traffic Safety Education, Vol. 19, No. 4, July 1972, pp. 16 and 26.
9. M. H. Jones. Measuring the Outcomes of Driver Training: University of Southern California Driver Performance Test. TRB, Transportation Research Record 629, 1977, pp. 63-67.

Publication of this paper sponsored by Committee on Driver Education.