Selection Tests—Dubious Aid in Driver Licensing

J. E. UHLANER and A. J. DRUCKER
U.S. Army Personnel Research Office

Tests developed for selection or screening of drivers are likely to be inappropriate for public licensing. Whereas selection tests seek to eliminate all but the best in a given applicant pool (a problem of interest to the Army, to commercial transportation concerns, etc.), licensing procedures concentrate on eliminating only the more obvious misfits. Tests that have been successful for Army driver selection include attitudinal, personality and/or adjustment measures developed through thorough empirical tryout, as well as tests of information. Psychophysical measures did not prove to be successful for selection purposes.

If license bureau officials made a decision to deny licenses to applicants scoring low on a typical driver selection battery validated against an accident criterion, it would be necessary to set the cutting score on the battery at a point where approximately 23 million drivers of 100 million, currently estimated to be licensed in the United States, would have to be taken off the road. This would achieve a reduction of fatalities and other accidents from 15 to 10 million per year. If officials arbitrarily removed 10 million drivers from the road on the basis of such a selection battery, only 2 million could be expected to be drivers likely to have accidents.

Limiting licensing in terms of personal limitations of the driver is regarded as a legitimate basic approach to reducing accidents. But a broader approach is needed because accidents occur as a result of multiple, complex causes or because they do not occur with the regularity and consistency needed for research. (An accident criterion could not be used in Army research for the latter reason. A criterion based on the observations and judgments of drivers, supervisors, and associate drivers was used instead.) Real-life simulation facilities are undoubtedly needed for the study of driver accidents as a means of deriving principles of engineering traffic, vehicles and roads and as a means of identifying driver limitations.

•MANY OF us would undoubtedly accept the contention that the driver is the most complex, baffling, and vulnerable of the factors that make up the driving process. This contention has given rise to the repeated query: why are people not given more complete examinations—physical, psychophysical, and psychological—before they are licensed to drive? Hundreds of tests have been developed and used in various settings; many have been used in driver research. Despite this enormous effort, few tests are in operational use, chiefly because of the following difficulties: (a) lack of evidence that many of these tests do an accurate job of screening out those likely to have

Paper sponsored by Committee on Road User Characteristics.

accidents; (b) lack of means of getting undisputed proof in terms of accidents; and (c) administrative and legislative barriers to using such tests for licensing.

Assuming that adequate tests were available for carrying out research, could the needed scientific proof of their effectiveness come from accident records? Turn any eager and inspired researcher loose with accident records and he will feel he should be able to test the worth of any theory for licensing purposes. However, accident records speak loudly but not clearly. So many variables contribute to accidents that no one variable, predominantly responsible though it may be in a specific accident, can measurably account for any significant proportion of the various accidents. The problem is multi-dimensional. Although traffic accidents occur all around us, paradoxically they occur too rarely for research. Research ideally needs replication of identical circumstances if responsible significant variables are to be isolated.

The driver on the road thwarts the advance of science because he can exercise common sense. He is adaptive in ways his vehicle is not. When at the wheel, he typically (though not always) compensates for his deficiencies. If he is color blind, he learns positions and shapes of traffic control signs. If his reflexes are slow, he tends to avoid getting into situations requiring fast stops or turns. He notes an appalling magnitude of accidents and drives defensively. Human factors scientists might well voice their complaint that the traffic accident "is not well enough organized for research."

So far the discussion has been concerned with the difficulties in carrying out research in this area, particularly research which might lead to more stringent testing and tighter licensing regulations. If stringent tests were used in licensing, the screams of rejected driver's license applicants would be heard in every state capitol. But considering the present state of the art, much of the wailing would, in our opinion, be justified.

It is important to make a differentiation at this point between predictive tests used for selection and tests used for licensing. The difference is crucial. In the case of selection and screening, management is interested in eliminating all but the best. In the licensing process, public officials concentrate on eliminating only the more obvious misfits. The Army is concerned with both the selection and the licensing problem, since many Army jobs require driving as an incidental duty. There are also many Army personnel whose primary duty is driving, e.g., Military Police and ammunition truck drivers.

The senior author devoted a number of years to the direction and conduct of research activities to develop devices for the selection and licensing of Army motor vehicle operators (8). The program has been large in scope but was justified because the U.S. Army is the largest user of motor vehicles in the world.

First, hundreds of existing tests were sifted since the literature appeared to be full of promising leads. Many of these leads were examined to serve as bases for research hypotheses, but most of these hypotheses were ultimately rejected. For many years an assumption existed among driver officials and researchers that visual and psychophysical measures were among the most effective predictors of efficient and safe driving. Close examination of the findings available at the start of the Army research did not bear out this hypothesis (1, 4). Admittedly, in the Army setting this hypothesis had less of a chance of being substantiated since military personnel in the classification stage of Army processing have already met certain minimum physical, visual, and psychophysical requirements for admission to military ranks. Hence such measures as field of vision, eye dominance, visual acuity, reaction time, depth perception, peripheral vision, auditory acuity, resistance to glare, and strength of grip could not be expected to differentiate as significantly among Army driver applicants as they do among civilian applicants. Further, there was little evidence that these measures had significant validity for the civilian population with respect to safe and efficient driving. Therefore, new measures were required to select further from the military manpower pool those Army personnel who would be safe and efficient drivers. Emphasis was placed on development of measures of driving information, emergency driver information, personality characteristics (in the form of likes and dislikes, attitudes, interests, and biographical information), and a variety of specially tailored psychological measures (10).

Almost 2,000 drivers were tested and their driving ability was examined by superiors and training NCO's (5). Six tests from a total of 22 were finally selected as the most predictive and arranged into the following operational selection batteries:

Battery I[1]

Driving Know-How Test—Knowledge of good driving practice.
Attention to Detail Test—A measure of perception requiring the rapid counting of the letter "C" interspersed among large numbers of the letter "O".
Army Self-Description Blank (Transport)[2]—A measure of personality and attitudinal factors such as interest, annoyances, likes, dislikes, preferences, driving and mechanical experience.
Age to 30 years.

Battery II[3]

Emergency Judgment Test—Knowledge of solutions to emergency driving problems.
Visual Judgment Test—Ability to match identical pairs of words as they are presented in progressively smaller type.
Two-Hand Coordination Test—A measure of eye-hand coordination.

It had been recognized at the outset of the research that a simple count of accidents would probably be inappropriate as a criterion of safe and efficient driving for three reasons:

1. During a single enlistment it would not be possible to obtain enough of a sample of the man's driving record to serve as a reliable index.
2. The distribution of accidents would be sharply curtailed in the Army because of the removal from driving duty of any driver who had had a second or third accident.
3. Driving conditions varied widely from motor pool to motor pool (2).

It was decided to employ a carefully constructed criterion based on the observations and judgments of drivers, supervisors, and associate drivers. An instrument was developed, including rating scales and a checklist. Drivers serving as examinees were rated on 11 experimental scales by an average of 4.8 supervisors and 12.5 associates. Of the 11 scales, four were chosen on the basis of (a) reliability coefficients, (b) correlation with an accident-responsibility index, (c) intercorrelation among the scales, and (d) results of a factor analysis of the intercorrelations (12).

[1] Validity coefficients on three samples of 331, 192 and 194 drivers ranged as follows: driving know-how, 0.31 to 0.41; attention to detail, 0.20 to 0.30; Army self-description blank, 0.18 to 0.41; equally weighted composite, 0.39 to 0.51.

[2] The Army Self-Description Blank for Transport may be of unusual interest and promise. The largest Army effort was devoted to the development of this 150-item test. Slightly more than half the items reflect the personality profile of the accident-prone individual. Other items are concerned with personality as demonstrated through attitudes toward the driving habits of others and self-estimates of driving habits and skills. An individual's judgment of another's driving may well reflect his own driving habits; the unsafe driver might check "Most drivers fail to stop completely at STOP signs." In Army research, driver experience such as knowledge of how to "soup-up" a car was found to be not unrelated to driver ability; hence, some "hot-rod" items might be included. For self-estimates of judgment, driving ability and reactions to frustrating situations, such items as "I can handle a car at high speeds" or "I am a careful driver driver" may be useful. Items attempting to measure past history of the driver may be appropriate, including difficulties he may have encountered with credit or disciplinary agencies—indicative of a negative complex in his total attitudinal behavior pattern as opposed to a clean slate indicating a positive complex (9).

[3] Validity coefficients on three samples of 331, 192 and 194 drivers ranged as follows: emergency judgment, 0.20 to 0.33; visual judgment, 0.15 to 0.34; two-hand coordination, 0.09 to 0.23; equally weighted composite, 0.24 to 0.28.

The four scales included in the criterion instrument were:

1. How often does he have near accidents?
2. How well does he react to sudden changes of traffic conditions?
3. How much does "temper" or "nerves" affect his driving?
4. How well does he know his own limitations—poor sight, slowness, lack of skill, etc. —and drive according to what he knows he can do?

The same raters were asked to indicate, for each of 105 descriptions of unsafe driving habits, how ratable (observable) the behavior was and how important it was to safe driving. The 15 statements adjudged most ratable and important were selected for the final checklist. The mean rating on the four scales received double weight and the mean number of checks received had unit weight in the composite criterion score. Sample checklist items included: "shows off when driving," "drives too fast for road conditions," and "follows other vehicles too closely" (11).

The predictive test battery finally developed had a reasonable amount of validity for the purpose of selection—in the range of 0.35 to 0.40 (8). It should be stressed that benefits from this validity can be achieved if the selection ratio is favorable, that is, when many more applicants for driving are presented for assignment than will ultimately be accepted. In the Army, this difficulty is only partially overcome by requiring that all replacement stream enlisted personnel processed through reception stations be administered Motor Vehicle Driver Selection Battery I. However, driving jobs in the Army do not get top priority comparable to combat, electronics and other jobs—perhaps they should not. So even here the selection ratio is not entirely favorable

Enlisted men not previously qualified on Driver Battery I, or officers and warrant officers who are to be considered for standard drivers' licenses, are tested at local installations by Army Motor Vehicle Driver Selection Battery II. A road test is an important part of licensing procedure and consists of a physical evaluation examination (visual acuity, field of vision, foot reaction time, and hearing) and of a driving performance test including manipulation of controls, practice run, depth perception test, check for emergency equipment, before-operation check, and location of instruments.

The three tests of Driver Battery I are scored as number of correct responses and sum obtained. The positive contribution of age to good driving is accounted for by adding to this sum a figure corresponding to two times the applicant's age in years to a maximum age of 30. The final figure is then converted to the Army Standard Score scale with a mean or numerical average of 100 and a standard deviation of 20. When these 20 points are added to and subtracted from the mean of 100, a framework is established into which about 68 percent of scores normally fall. Standard scores, as the name implies, help establish standard interpretations of test performance. At reception stations, a standard score of 90 on Battery I is passing and serves as a screen for further testing (road testing) and licensing for driver vacancies. A standard score of 90 is that score achieved by 65 to 70 percent of all applicants when standardized on a sample which roughly approximates the applicant population.

The scores of Battery II are the number of correct answers converted to Army Standard Score units and averaged. A passing score on Battery II is 80 or may be placed higher for a greater degree of driver judgment or responsibility in selected assignments, as in that of Military Policeman or Investigator.

A score of 70 on the road test is the final requirement for licensing. However, weaknesses revealed on any portion of the physical evaluation or the road test are brought to the attention of the examinee as a basis for further practice or training or for his awareness (so that he can allow for his deficiencies when driving), whether the road test is successful or not. The physical evaluation standards include 20/30 acuity in each eye, a lateral range of 75° on each side of the focus line, foot reaction time to and including 0.60 seconds, and ability to hear the whispered voice at 15 feet. Thus, to obtain an Army Motor Vehicle (Transport) license, a man must pass Driver Battery I at the reception station or Driver Battery II at his local installation, take a physical evaluation test, and pass the road test.

One problem which concerned Army researchers may be of interest. Faced with the requirement of obtaining qualified drivers in foreign countries, the U.S. Army considered appropriate tests for selecting indigenous personnel as drivers, particularly in countries where the motor vehicle is practically a rarity (3). One difficulty is language. Another, perhaps even more serious, is the cross-cultural gap which is not bridged automatically with direct translations—a fact which Army human factors researchers had learned in connection with other research programs. The approach was to construct a battery of tests appropriate for non-English speaking nationals. Several tests were developed with pantomine administration instructions, including types used in the regular Army Driver Battery—attention to detail, two-hand coordination, emergency judgment, and driving know-how. Included were tests of mechanical principles, tool usage, driving concepts, and ability to perceive change in detail of abstract patterns of automotive equipment. A tryout of the regular test battery along with the new tests indicated the feasibility of non-language tests, and such a battery is now available for use when necessary.

In another special study, the U.S. Army Personnel Research Office considered whether differential requirements should be stipulated for drivers of light vs heavy vehicles (6). The tests developed for the selection of Army drivers of wheeled vehicles of any kind did not show practicality for differentiating drivers with good potential for vehicles of differing weights, although one set of tests isolated showed slightly more validity for heavy vehicles.

It seems important to offer some discussion of the significance of selection tests with specified validity coefficients as they relate to possible use in public licensing of drivers. In 1962, there were nearly 100 million persons in the United States licensed to drive vehicles. About 15 percent or 15 million were involved in fatal and other accidents. How many drivers would have to be taken off the road to reduce the total number of accidents to 10 million per year?

We think that a liberal estimate of validity of a good selection battery for drivers is 0.35 for drivers of Army vehicles, using a rating criterion. But because a rating criterion does not necessarily coincide with variance of actual incidence of accidents, we reduced our estimate to a validity coefficient of 0.20. Using this coefficient, public officials would have to take 23 million drivers off the road to reduce the number of accidents to 10 million per year (Fig. 1). Further, the cost to the public would be a loss of 18 million good drivers for the benefit of removing 5 million poor drivers (see Appendix for statistical methods used in estimating these and later figures).

If, through the expenditure of funds for additional research effort, we could raise the validity coefficient to 0.35, public officials could reduce the number taken off the

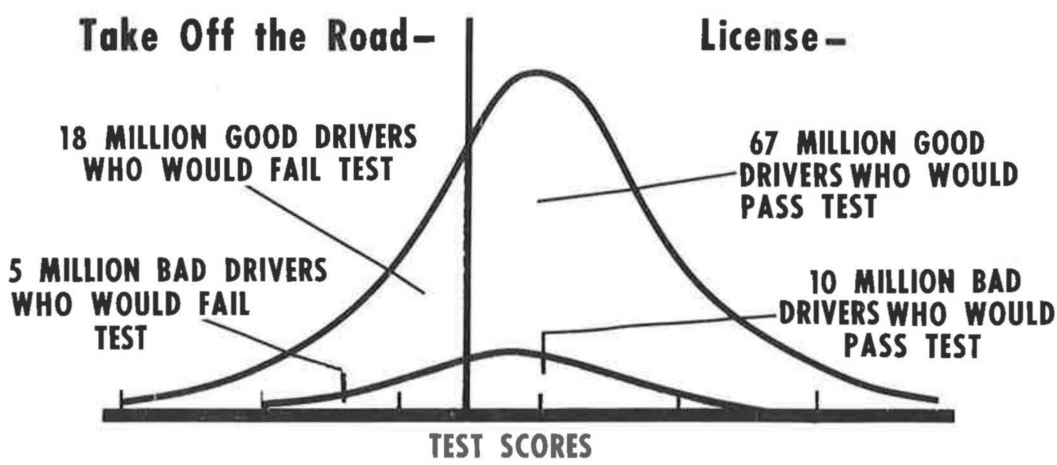

Figure 1. Impact of selection battery for licensing using validity coefficient of 0.20.

road to 18 million to achieve the reduction to 10 million accidents per year (Fig. 2). The loss this time would be 13 million good drivers.

What would be the impact on accident reduction if public officials would be willing to remove 10 million drivers? Using a validity coefficient of 0.20, we might expect only 2 percent of the 10 million to be bad drivers (Fig. 3). With a validity coefficient of 0.35, we might expect 3 percent of the 10 million to be bad drivers (Fig. 4). In each case, the increase in validity results in only slight improvement for the research money invested.

But consider a case of much higher validity. Table 1 provides the answer where 10 percent is retained as the point of cut or disqualification. If a predictive validity coefficient of 0.90 could be achieved, about nine-tenths of the drivers removed would be

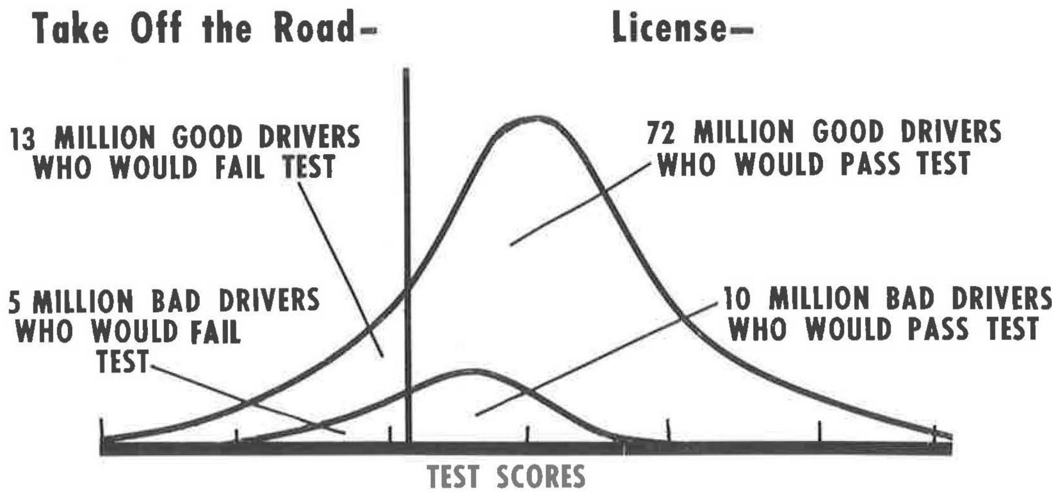

Figure 2. Impact of selection battery for licensing using validity coefficient of 0.35.

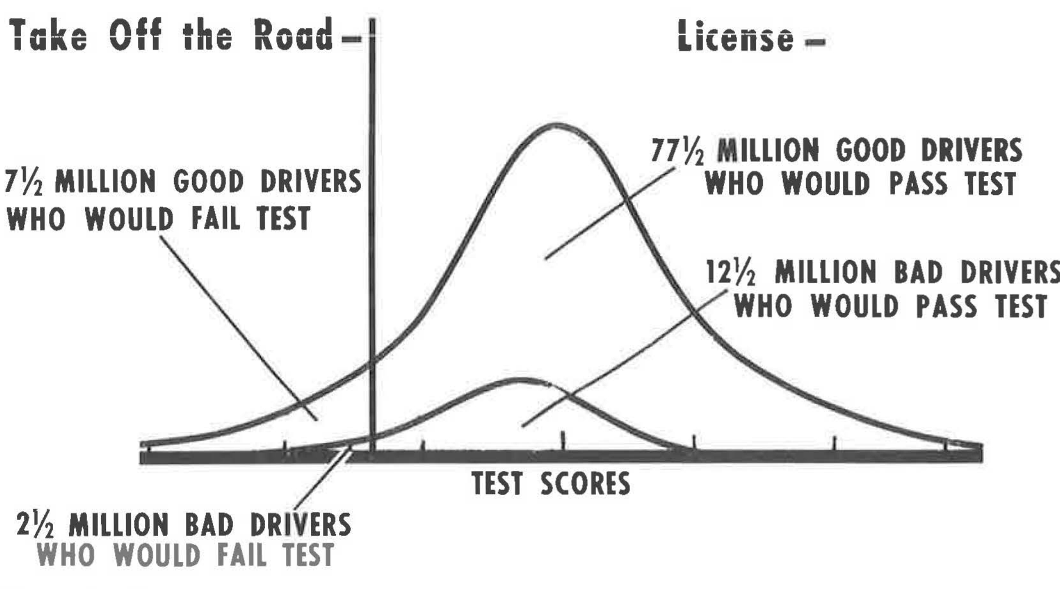

Figure 3. Impact of not licensing lowest 10 percent, with validity coefficient of 0.20.

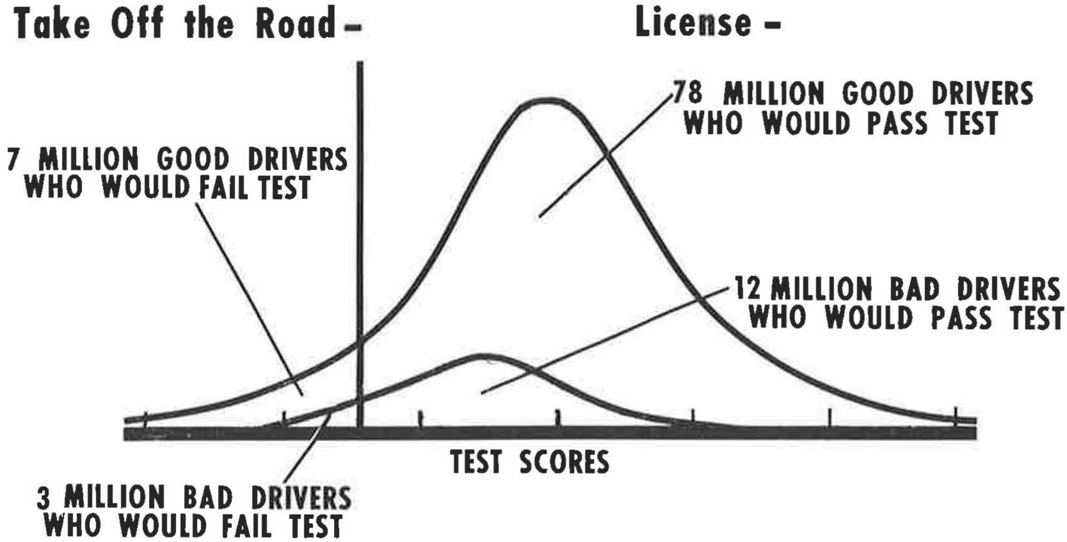

Figure 4. Impact of not licensing lowest 10 percent, with validity coefficient of 0.35.

TABLE 1

EFFECT ON DRIVER ACCIDENT REDUCTION OF REMOVING BOTTOM 10 PERCENT USING SELECTION TESTS FOR LICENSING

Validity Coefficient	Total Removed (millions)		Total Licensed (millions)	
	Bad Drivers	Good Drivers	Bad Drivers	Good Drivers
0.10	1.9	8.1	13.1	76.9
0.20	2.4	7.6	12.6	77.4
0.30	3.0	7.0	12.0	78.0
0.40	3.5	6.5	11.5	78.5
0.50	4.2	5.8	10.8	79.2
0.60	5.0	5.0	10.0	80.0
0.70	5.9	4.1	9.1	80.9
0.80	7.1	2.9	7.9	82.1
0.90	9.1	0.9	5.9	84.1

bad drivers and only one-tenth good drivers, and the bad drivers licensed would be reduced to about 6 percent. Similarly (Table 2) for a validity coefficient of 0.90, with a goal of reduction of the annual accident rate from 15 million to 10 million, virtually all good drivers tested in applying for licenses would receive them. Of course, these examples remain highly theoretical, since no immediate prospects exist for raising validity coefficients beyond present levels.

We may have simplified the picture a bit in that we have not taken into account the supposition that with fewer cars on the road, the progression of reduced accidents would not necessarily be a straight line, but might accelerate in curvilinear fashion. Our main purpose, however, is to illustrate why the present state of the art in driver selection research does not yield a dramatic solution to the problem of reducing our national motor vehicle accident rates when such selection devices are employed in the practical setting of general licensing.

One reason for the ineffectiveness of selection tests applied to a licensing situation is the administrative necessity to leave the point of cut at a low level. Nevertheless, the reader might be interested in learning how selection tests having validity coefficients

TABLE 2

EFFECT ON DRIVER ACCIDENT REDUCTION OF USING SELECTION TESTS FOR LICENSING[a]
(In millions)

Coefficient	Total Removed (millions)		Total Licensed (millions)	
	Bad Drivers	Good Drivers	Bad Drivers	Good Drivers
0.10	5.0	23.0	10.0	62.0
0.20	5.0	18.2	10.0	66.8
0.30	5.0	14.0	10.0	71.0
0.40	5.0	11.4	10.0	73.6
0.50	5.0	7.5	10.0	77.5
0.60	5.0	5.1	10.0	79.9
0.70	5.0	3.2	10.0	81.8
0.80	5.0	1.7	10.0	83.3
0.90	5.0	0.6	10.0	84.4

[a]To reduce accident rate from 15 million to 10 million.

as low as 0.35 or even 0.20 can be of value for selection purposes. Figure 5 and 6 illustrate this phenomenon graphically. As the vertical bar is moved to the right indicating a progressively higher cutting score on the test battery, the 100 million decreases, of course, but the proportion of poor drivers among those selected decreases more rapidly than the proportion of good drivers. Such a procedure requires that the number of applicants far exceed the number to be selected and be practical in only very limited commercial situations. Tables 3 and 4 give the values for 10, 25, 50, 75 and 90 percent points of cut. Maximally effective selection is achieved for a validity coefficient of 0.35 when only the top 10 percent is selected—9.7 percent accident-free drivers vs 0.3 percent accident drivers!

In summary, our research experience in the Army with selection tests and selection batteries is this:

1. Use of selection procedures in public licensing, at least with the types of variables now generally in use, can make only a slight contribution to the accident reduction problem.

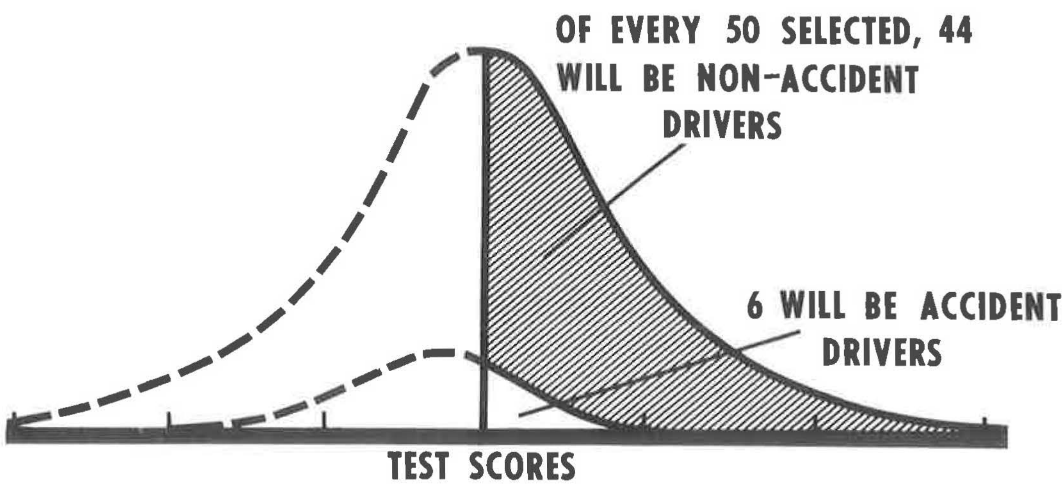

Figure 5. Influence of cutting off bottom 50 percent of selection applicants, given validity coefficient of 0.20.

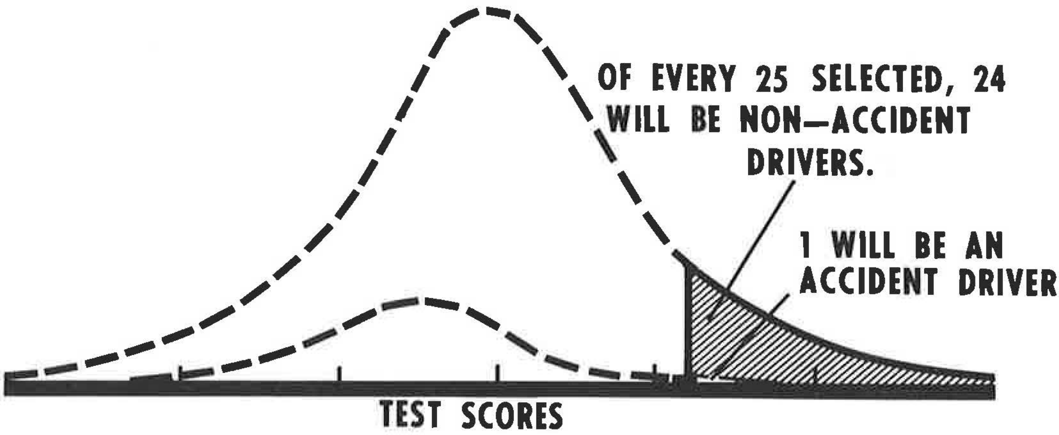

Figure 6. Influence of cutting off bottom 75 percent of selection applicants, given validity coefficient of 0.35.

TABLE 3

IMPACT ON ACCIDENT REDUCTION OF REJECTING VARIOUS PROPORTIONS OF DRIVER APPLICANTS, VALIDITY COEFFICIENT OF 0.20

Selection Ratio[a]	Percent Rejected		Percent Selected	
	Bad Drivers	Good Drivers	Bad Drivers	Good Drivers
10	2.4	7.6	12.6	77.4
25	5.3	19.7	9.7	65.3
50	9.4	40.6	5.6	44.4
75	12.6	62.4	2.4	22.6
90	14.2	75.8	0.8	9.2

[a]Percent to be eliminated.

TABLE 4

IMPACT ON ACCIDENT REDUCTION OF REJECTING VARIOUS PROPORTIONS OF DRIVER APPLICANTS, VALIDITY COEFFICIENT OF 0.35

Selection Ratio[a]	Percent Rejected		Percent Selected	
	Bad Drivers	Good Drivers	Bad Drivers	Good Drivers
10	3.0	7.0	12.0	78.0
25	6.6	18.4	8.4	66.6
50	11.0	39.0	4.0	46.0
75	13.8	61.2	1.2	23.8
90	14.7	75.3	0.3	9.7

[a]Percent to be eliminated.

2. A reasonable amount of success has been possible in the use of selection batteries to assist in selecting only the best drivers in terms of the likelihood of fewer accidents occurring. To generalize to commercial driving, if relatively few are to be selected from among many who apply, selection devices will contribute to more efficient and safe motor vehicle operation.

3. Of selection devices thus far developed and submitted to research evaluation, attitudinal factors, particularly as reflected by personality and adjustment measures reported in this paper, probably can make a significant contribution. But truly effective measuring devices of this nature would have to be developed on the basis of empirical data obtained. (Many of the research studies conducted today in the driver research area reveal background and personality results which are next to uninterpretable because the studies deal with extreme cases only—those with many accidents and those with striking records of absence of accidents—with the bulk of the drivers in the normal range being omitted.)

4. In general, psychophysical measures (visual and auditory skills and capacities, physical coordination and reaction time) make only minor contributions to predicting driver performance. On the other hand, a factor such as age usually is significant in that younger people tend to be more identifiable in the negative complex of driving behavior.

Licensing in terms of the personal limitations of the driver is still a legitimate basic approach to reducing accidents. But a broader approach is obviously needed for the traffic accident usually occurs not as a result of a single variable—inattentiveness because of fatigue or preoccupation, slippery roads, or insufficient light—but as a result of a complex of variables. Indeed, one of the encouraging signs of progress attributable to driving safety researchers is their recent success in reducing the total problem to manageable proportions. Just as the military man and the weapon or machine he serves and the enviroment in which he performs his assigned duties are all viewed as a man-machine or man-weapons system, so should the driving process be considered a system. Viewed this way, malfunction of the driver system can occur because of (a) poorly designed and maintained vehicles, (b) poor roads and poorly controlled traffic patterns, and (c) poor driving.

We believe ultimate reduction of accidents is likely to come about through more effective human engineering of the automobile, the road, and the traffic system, as well as through greater effort in understanding the driver process. Particularly needed is a better understanding of relationships involved in various situational behaviors (psychological functioning in driving both at night and in daylight on turnpikes, in rural areas, and in the city).

The research approach dictates highly sophisticated simulation facilities and should be directed toward the alternate outcomes of educating the potential driver to difficulties inherent in a variety of conditions or limiting the situations in which he may be permitted to drive.

In conclusion, if public officials are inclined to shrink away from action which would eliminate millions of drivers from the road to reduce the national accident rate, then it may be profitable to embark on the so-called systems approach of driving research. This would be essentially a reexamination of the total problem to consider man-vehicle-road-traffic and to derive principles of engineering traffic, vehicles, roads, and identifiable driver limitations.

ACKNOWLEDGMENT

The authors are obligated to John J. Mellinger, U.S. Army Personnel Research Office, who provided consultation on statistical methods presented in the Appendix.

REFERENCES

1. DeSilva, H.R., and Robinson P. Some Psychological Factors in Accident Repeater Drivers. Jour. Abn. Soc. Psychol., Vol. 34, pp. 124-128, 1939.
2. Edgerton, H.A. Personnel Management Factors in Vehicle Safety. Tech. Res. Rept. 911, TAGO, May 1951.
3. Flanagan, J.C., and Loveland, E.H. Driver Selection Tests for Non-English Speaking Nationals. PRB Tech. Res. Rept. 1074, TAGO, Aug. 1953.
4. Goldstein, L.G., Van Steenberg, N.J., and Birnbaum, A.H. Evaluation of Instruments to Select Safe Drivers. PRS Rept. 962, TAGO, July 1952.

5. Lauer, A.R. Aptitude Tests for Army Motor Vehicle Operators. PRS Rept. 981, TAGO, Oct. 1952.
6. Lauer, A.R., and Vavra, L.R. Selection Tests for Army Operators of Heavy and Light Motor Vehicles. PRB Tech. Res. Rept. 1081, TAGO, Dec. 1953.
7. Staff, Personnel Research Section, TAGO. Personnel Research in the Army: VI. The Selection of Truck Drivers. Psychol. Bull. Vol. 40, pp. 499-508, 1943.
8. Uhlaner, J.E. Tests for Selecting Drivers. Presented at Eastern Psychological Assoc. Meeting, March 1956.
9. Uhlaner, J.E., Van Steenberg, N.J., Brogden, H.E., and Sass, Tauba. Construction of the Drivers Self-Description Blank. PRS Res. Note 51-33, TAGO, Dec. 1951.
10. Uhlaner, J.E., Van Steenberg, N.J., and Goldstein, L.G. The Construction of Experimental Group Tests for the Prediction of Safe Driving. PRS Res. Note 51-40, TAGO, 1951.
11. Uhlaner, J.E., Van Steenberg, N.J., and Goldstein, L.G. The Development of Criteria of Safe Driving for the Individual, Part I: The Driving Habit Check List. PRS Res. Note 51-13, TAGO, 1951.
12. Uhlaner, J.E., Van Steenberg, N.J., and Goldstein, L.G. Development of Criteria of Safe Motor Vehicle Operation. Highway Research Board Bull. 60, pp. 1-16, 1952.

Appendix

STATISTICAL METHODS FOR ESTIMATING EFFECT OF SELECTION TESTS ON ACCIDENT REDUCTION

Question 1

Assuming a biserial correlation of 0.20 between a driving ability test and a criterion (no accidents vs some accidents), and assuming criterion frequencies of 85 million no-accident drivers and 15 million accident drivers, how many drivers with low test scores should be eliminated in order to reduce the 15 million accident group to 10 million?

Figure 7 conceptualizes the problem in which a continuous varaible x corresponds to scores on the driving ability test, a dichotomous variable y corresponds to criterion

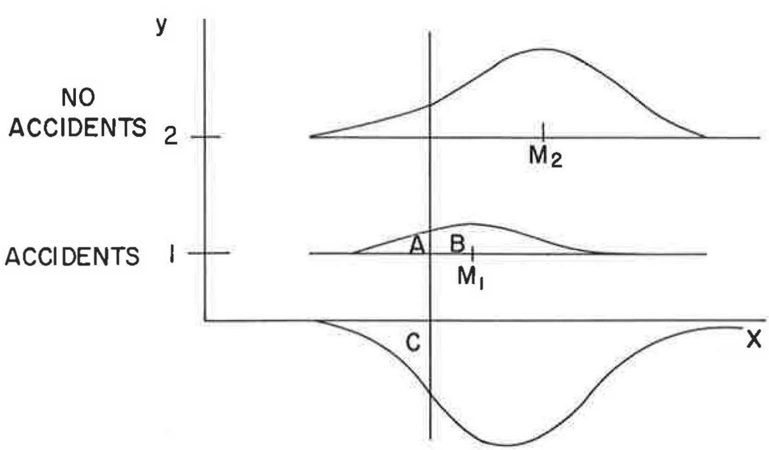

Figure 7. Method for determination of cutting score for accident vs no-accident drivers.

performance, M_2 is the mean x score of the no-accident drivers, and M_1 is the mean x score of the accident drivers. The numbers of people in the criterion groups are 85 million and 15 million, respectively. Scores of 2 and 1 can be arbitrarily assigned to the criterion groups. The vertical cutting line divides the accident group, or No. 1 distribution, into areas A and B, with 5 and 10 million people, respectively. The problem is to determine how many people are contained in area C. The methods described below assume normality of both marginal and conditional x distributions. This assumption is reasonable if the correlation between x and y is low.

Using the biserial correlation formula

$$r_b = \frac{(M_2 - M_1)\, pq}{Z\sigma x} \tag{1}$$

the assumptions provide sufficient information to solve for the difference $M_2 - M_1$. A second equation involving M_2 and M_1 can be written by making use of the fact that their weighted sum is zero. Both equations are expressed in terms of normal deviates of the x distribution. These two equations are $M_2 - M_1 = 0.366$ and $0.85M_2 + 0.15M_1 = 0$. Solving for M_1, we obtain $M_1 = -0.311$. This number describes the extent to which the No. 1 distribution is shifted to the left of the marginal x distribution.

The cutting line divides the No. 1 distribution into two parts, A and B. The corresponding normal deviate, in terms of the No. 1 distribution, is -0.43. To express this number in terms of the standard deviation of the x distribution, we must take into account that the No. 1 distribution has a standard deviation slightly smaller than the x distribution. The No. 1 standard deviation is described in the equation

$$\sigma_1 = \sqrt{1 - r^2} = \sqrt{1 - 0.20^2} \tag{2}$$

for which the solution is 0.980. The normal deviate of -0.43 multiplied by the standard deviation of 0.980 gives -0.421, which is expressed in terms of the standard deviation of the x distribution. This number, when added to the mean of the No. 1 distribution ($M_1 = -0.311$), gives a final result of -0.732. This number describes the location of the cutting line in terms of normal deviates of the x distribution. Using tables of the normal curve, we learn that the area below the cutting line (area C) is 0.232, or 23.2 million people.

The conclusion is that roughly 23 million drivers would have to be eliminated to remove 5 million accident drivers. Of these 23 million, 18 million constitute no-accident drivers.

Question 2

Assume a correlation of 0.35 instead of the correlation of 0.20 used previously. Using the same method we get the following results: $M_1 = -0.544$; $\sigma_1 = 0.875$; $-0.43 \times 0.875 = -0.376$; $-0.376 + -0.544 = -0.920$.

The final number, -0.920, describes the location of the cuttling line with respect to the x distribution. The area below this normal deviate is 0.179, or 17.9 million people. Thus, to remove 5 million accident drivers, a total of about 18 million drivers would have to be eliminated.

Question 3

Assuming a correlation of 0.20, how many accident drivers are removed by eliminating the 10 million drivers who scored lowest on the driving ability test?

The method of solution here is the reverse of that of Question 1. We are given area C and wish to solve for area A. The normal deviate corresponding to the bottom 10 percent of the x distribution is -1.281. Subtracting the mean of the No. 1 distribution

(-0.311) gives -0.970; when this number is converted to normal deviates of the No. 1 distribution by dividing by 0.980, the resulting normal deviate is -0.990. The area below this normal deviate is 0.161, which in the present case corresponds to 2.42 million people. In other words, removing the 10 million drivers with lowest scores on the driver ability test eliminates about 2.4 million accident drivers. The ratio of total drivers to some-accident drivers, about four, is roughly the same here and in Question 1.

Question 4

Assume a correlation of 0.35, instead of the correlation of 0.20 used in Question 3. From the same initial normal deviate of -1.281, the mean of the No. 1 distribution (-0.544) is subtracted; the result of -0.737 is converted to normal deviates of the No. 1 distribution by dividing by 0.875, giving a final normal deviate of -0.842. The area below this deviate is 0.200, corresponding to 3 million people. Thus, if the bottom 10 percent of all drivers are removed, 3 million accident drivers are eliminated.