

Reproduced by

DOCUMENT SERVICE CENTER

ARMED SERVICES TECHNICAL INFORMATION AGENCY

U. S. BUILDING, DAYTON, 2, OHIO

REEL-C

540

A. T. I.

162129

"NOTICE: When Government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the U.S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto."

UNCLASSIFIED

UNCLASSIFIED

ATI 162 129

(Copies obtainable from ASTIA-DSC)

Adjutant General's Office, Personnel Research Section, Wash., D.C.
(CRS Report 935)

The Development of Criteria of Safe Driving for the Individual

Uhlaner, Julius E.; Van Steenberg, Neil J.; Goldstein, Leon G.
4 April 52 1pp. tables, graphs

Accidents - Prevention
Vehicles, Passenger
Aptitude tests

Military Operations (24)
Personnel and Training (3)

UNCLASSIFIED

Army Project Number
29565100

PJ 6501-01

PRS REPORT 935

THE DEVELOPMENT OF CRITERIA OF SAFE DRIVING FOR THE INDIVIDUAL

Julius E. Uhlaner, Neil J. Van Steenberg, Leon G. Goldstein

Approved by Julius E. Uhlaner,
Program Coordinator

4 April 1952

PRS reports are primarily technical. While conclusions affecting military policy or operations may appear in them, they are not intended as a basis for official action. Findings and conclusions contained in PRS reports are intended to guide the conduct of further research. When research findings suggest recommendations for administrative action, such recommendations are made separately to the appropriate military agency.

BRIEF

MEASURING THE SAFETY OF MOTOR POOL DRIVERS (Based on PRS Report 935)

STATEMENT OF THE PROBLEM

In order to determine the effectiveness of tests in selecting safe drivers for Army motor pools, it is necessary to have a yardstick of safe driving in the Army situation. The purpose of this project was to design such a yardstick using the most meaningful and dependable information obtainable - whether gained from the drivers' accident records and road tests or from observations made by their fellow-drivers and supervisors.

RESULTS

1. Both accident records and road tests were considered as bases for this measure.
 - a. An extensive survey of the literature revealed that civilian vehicle accident records were unstable over a period of time. Furthermore, accident records of Army drivers were found to cover only short periods of driving and to vary in meaning because of differences in driving conditions.
 - b. Road tests had the disadvantage of measuring drivers' performances in the presence of examiners. Such performances may be quite different from driving behavior under ordinary conditions.
2. Pooled observations and ratings made on drivers by their supervisors and fellow-drivers were finally chosen as the basis for the measure. To convert these into meaningful and comparable scores of safe driving, 11 experimental rating scales and 105 experimental check list items were devised. These scales and items were given field trial and the resulting data analyzed statistically to determine which scales and items were most stable, most meaningful, and least overlapping.
3. On the basis of agreement between ratings and an Accident Responsibility Index developed in this program, the most effective raters were found to be fellow-drivers with Aptitude Area I scores above 90 and supervisors.

CONCLUSIONS

Pooled observations and judgments on drivers by their supervisors and fellow-drivers were more stable and meaningful than either existing accident records or road tests. The best judgments were those made on four rating scales and 15 check list items by fellow-drivers who had Aptitude Area I scores above 90 and supervisors.

WORK SUMMARY

An extensive survey was made of the technical literature on accident records. Anonymous ratings for 200 Army drivers on the 11 experimental rating scales and evaluations of the 105 experimental check list items were obtained from 158 drivers and 35 supervisors in seven motor pools of the First Army area during July 1960. Results were analyzed using correlational and factor analysis methods.

THE DEVELOPMENT OF CRITERIA OF SAFE DRIVING FOR THE INDIVIDUAL

CONTENTS

	Page
NATURE OF THE PROBLEM	
Inadequacy of accident records	1
Civilian studies	1
Conditions peculiar to the Army situation	3
Inadequacy of road tests	3
Specific approach of this project	4
DEVELOPMENT OF RATING SCALES	
Construction of experimental scales	4
Administration of experimental scales	5
Analysis of results	5
Reliability of experimental scales	5
Correlation of mean ratings with Accident Responsibility Index	6
Intercorrelations among the scales and factor analysis of the intercorrelation matrices	8
Selection of the final scales	10
DEVELOPMENT OF DRIVING HABIT CHECK LIST	
Construction of experimental items	10
Selection of final items	13
Relation between selected check list items and experimental rating scales	13
FINAL COMPOSITE CRITERION	
The final criterion instrument	15
Scoring of the final criterion instrument	15
Selection of raters for final criterion ratings	15
REFERENCES	19

TABLES

Table	Page
1. Accidents of general drivers in Connecticut in years 1934-36 compared with those of the same drivers in years 1931-33.	1
2. Product-moment correlations between accidents in separate years of exposure.	2
3. Means and standard deviations of mean ratings on 11 driver rating scales, and the reliabilities of the 11 scales.	6
4. Distribution of Accident Responsibility Index.	7
5. Correlations between mean ratings and Accident Responsibility Index.	7
6a. Intercorrelations among mean ratings by supervisors on 11 driver rating scales.	8
6b. Intercorrelations among mean ratings by associates on 11 driver rating scales.	9
7. Factor loadings on orthogonal axes derived from intercorrelation matrices of 11 driver rating scales.	9
8. Twenty-one behavior statements rated "very important" and "I know I can rate on this" by 49% or more of drivers and supervisors.	14
9. Assignments of selected check list items to rating scale continua.	15
10. Correlations between ratings by supervisors and ratings by different grades of associates.	16
11. Correlations of mean ratings with Accident Responsibility Index: Group A with mean AAI below 90, Group B with mean AAI above 90.	17

FIGURES

Figure	Page
1. Distributions of mean ratings of 189 drivers rated by associates.	11
2. Distributions of mean ratings of 121 drivers rated by supervisors.	12

THE DEVELOPMENT OF CRITERIA OF SAFE DRIVING FOR THE INDIVIDUAL

NATURE OF THE PROBLEM

The continuing cost of motor vehicle accidents in the three Armed Services led to joint action which culminated in 1949 in the establishment of a broad program of research to investigate psychological factors involved in safe operation.^{1/} Two projects were established for the development of selection tests, and this project was established for the development of an acceptable criterion against which to evaluate these tests.

INADEQUACY OF ACCIDENT RECORDS

Civilian Studies. Although most of the studies on safe driving in the civilian situation have used accident records as the criterion, only a very few have attempted to evaluate the stability of accident records. The most comprehensive data on the subject are furnished by the Connecticut study of the Bureau of Public Roads(3) which used the 6-year records (1931-36) of over 29,000 licensed drivers. In Table 1 the distribution of accidents in the second 3-year period is compared with that of the same drivers in the first 3-year period.

Table 1. Accidents of general drivers in Connecticut in years 1934-36 compared with those of the same drivers in years 1931-33 (from Bureau of Public Roads Study)

Accidents per operator, 1934-36	Accidents occurring to same operators, 1931-33					Total number of operators
	0	1	2	3	4	
0	23881 (91%)	2386 (83%)	275 (77%)	22 (71%)	5 (50%)	26369 (90%)
1	2117 (8%)	419 (15%)	64 (18%)	5 (16%)	4 (40%)	2609 (9%)
2	242 (.9%)	57 (2%)	12 (3%)	2 (6%)	0 (0%)	313 (1%)
3	17 (.6%)	9 (.3%)	5 (1%)	2 (6%)	1 (10%)	34 (.1%)
4	2 (.0%)	3 (.1%)	1 (.3%)	0 (0%)	0 (0%)	6 (.02%)
Totals	26259 (99%)	2874 (9.7%)	357 (1.2%)	31 (.1%)	10 (.03%)	29531 (100%)

This table shows a definite tendency of those who were accident-free in the first period to be accident-free in the second period, and a progressive likelihood of those who had accidents in the first period to have accidents in the second period. However, the correlation (tetrachoric) of accident experience in the two periods comes to .24.

Farme and Chambers(3) correlated accident experience in separate years for four groups of British public transportation drivers. These correlations are shown in Table 2.

^{1/} Program Plan, PR-4123: Research on Motor Vehicle Accident Problem, 19 Aug 49.

^{2/} Percentages inserted by present authors.

Table 2. Product-moment correlations between accidents in separate years of exposure (from Farmer and Chambers)

Correlation between accidents in years:	Group A 166 Bus Dr.	Group B 396 Bus Dr.	Group C 86 Bus and Trolley Dr.	Group D 67 Trolley Dr.
1 + 2	.298	.182	.235	.071
1 + 3	.235	---	.063	.058
1 + 4	.177	---	.281	.127
1 + 5	.274	---	---	---
2 + 3	.328	---	.078	.225
2 + 4	.176	---	.195	.251
2 + 5	.265	---	---	---
3 + 4	.212	---	.016	.296
3 + 5	.273	---	---	---
4 + 5	.224	---	---	---

For the 166 bus drivers in Group A the correlations between accidents in the first year and accidents in increasing subsequent periods were as follows:

Between 1st and 2nd year	$r = .298$
Between 1st and 2-3 years	$r = .327$
Between 1st and 2-4 years	$r = .339$
Between 1st and 2-5 years	$r = .375$

Brown and Ghiselli(2) estimated the reliability coefficients from correlations between the number of accidents on the odd and even months over a period of 18 months, corrected by the Spearman-Brown formula. For 59 trolley car motormen in California the estimates are as follows for different types of accidents:

Collision with pedestrians	$r = .46$
Collision with trolley cars	$r = .19$
Collision with motor vehicles	$r = .42$
All collision accidents	$r = .42$

Bransford(1) correlated accident frequency during the year and a half after administration of driver tests against accident frequency during a variable period prior to testing. The correlation between accident rates, for a group of 481 drivers in D. C., was .184.

Slocumbe(5) correlated yearly accident rates of 260 motormen of the Boston Elevated Railroad over a period of four years. The correlations between different yearly periods were as follows:

1st and 2nd year	$r = .51$
1st and 3rd year	$r = .43$
1st and 4th year	$r = .38$
2nd and 3rd year	$r = .41$
2nd and 4th year	$r = .38$
3rd and 4th year	$r = .45$

Conditions Peculiar to the Army Situation. It was hardly to be expected that accident records in the Army situation would be more stable than those of these civilian studies. On the contrary, they might well be less stable, for the following reasons:

- (1) The average mileage of Army drivers per year is approximately 12,000 to 15,000 miles. Even if a driver were kept on duty at the same pool during a full enlistment and his records were therefore available, this extent of exposure would hardly be sufficient to yield a reliable measure of safety of driving behavior.
- (2) The distribution of accidents is curtailed because of the policy of removing a driver from driving duty when he has had a second or third accident.
- (3) Driving conditions from pool to pool vary greatly in terms of vehicles used, mission, supervision, climate, terrain, density of traffic, night versus day driving, etc. Accident rates would thus be contaminated with uncontrolled variables.
- (4) Civilian records of recruits' driving behavior prior to their entry into the Army are not available.

INADEQUACY OF ROAD TESTS

Previous studies of the Personnel Research Section(6) furnish considerable data on the reliability of road tests. Four kinds of estimates of reliability were made.

- (1) The relationship between separate check-list items and general ratings on the road test ($N = 1717$) ranged between $r = .22$ and $r = .57$.
- (2) The split-half reliability of the Road Test Check List, using the Spearman-Brown formula, was computed as $r = .82$ ($N = 155$).
- (3) The correlation of scores on the Road Test Check List given by different examiners at different times (same 155 cases) was $r = .53$.
- (4) The reliability of the general ratings on the road test was computed by correlating two series of ratings of 127 men made on the same day by several specially trained examiners. This yielded a coefficient of $r = .72$.

Although these reliabilities are generally higher than those of accident records, the road test suffers from one serious objection as a measure of safe driving behavior, regardless of its reliability. Driving behavior during such a test, under the surveillance of one whom the driver recognizes as an examiner, may be expected to be different from driving behavior under ordinary conditions.

SPECIFIC APPROACH OF THIS PROJECT

In the light of the shortcomings of both accident records and road tests as criteria for the evaluation of instruments for the selection of safe drivers, the decision was made to explore the possibility of assessing the driving behavior of Army drivers on the basis of the observations and pooled judgments of their supervisors and associates. In view of the practical considerations of administration, and in view of the generally high intercorrelations obtained among rating scales and check lists, it was decided to develop a criterion instrument of about four simple rating scales and a driving habit check list of about 15 items.

DEVELOPMENT OF RATING SCALES

CONSTRUCTION OF EXPERIMENTAL SCALES

Eleven experimental scales were constructed from which the final four were to be selected. Eight aspects of driving behavior were postulated and a 15-point rating scale was designed to measure each. Two additional scales were designed to be used as psychological suppressors and another scale was designed for an over-all safe-unsafe rating. The lead questions of the 11 scales are given below.

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 - like poor eyes, slowness, lack of skill, etc. - and drive according to what he knows he can do?
- *5. How safe a driver is he?
6. What is his attitude toward safety when he drives?
7. How well does he keep his mind on his driving?
- *8. How well do you like him?
9. How skillful is he in handling a vehicle?
- *10. How does he rate on appearance and military bearing?
11. How well does he take care of his vehicle?

*Designed to be used as psychological suppressors.

**Designed to obtain an over-all safe-unsafe rating.

Each scale was divided into 5 sections (3 scale points each) with a verbal definition of each section printed therein. On ten of the 11 scales cartoons designed for the purpose of illustrating the intent of the lead question were printed at the "good" and "bad" ends of the continuum. The 11 scales were printed in booklet form, preceded by a practice rating scale (FRT 2057). The booklet was arranged so that the rater could not see, as he worked, what ratings he had given on the previous scale.

ADMINISTRATION OF EXPERIMENTAL SCALES

A trial run of this booklet was conducted during July 1950 in seven motor pools in the First Army Area. The motor pools visited were at Fort Jay, N.Y., West Point, N.Y., Manhattan, N.Y., and Fort Dix, N.J. Rosters of not more than 20 drivers each were drawn up in such a manner that familiarity with one another's driving behavior was maximal. In rating sessions, conducted with drivers and with their supervisors separately, a total of 200 drivers were each rated by 2-10 (mean of 4.8) supervisors, and by 5-26 (mean of 12.5) associate drivers. A total of 23 sessions were conducted. All ratings were anonymous.

ANALYSIS OF RESULTS

In order to select the four scales which would best measure safe driving behavior, the following possible bases of selection were investigated:

- (1) Reliabilities of the 11 scales
- (2) Correlations of mean ratings on the scales with an index of accident responsibility
- (3) The results of a factor analysis of the intercorrelations among the ratings which was intended to identify those scales with high loadings on that orthogonal factor (or factors) which represents the variance of most of the scales and have low loadings on the orthogonal factor identified with the suppressor scales.

Reliability of the Experimental Scales. The reliability of each scale was estimated by means of a modification of the Horst formula(4):

$$rel = 1 - \frac{\sum_{i=1}^N \frac{n_i^2 - 1}{n_i - 1}}{\sum_{i=1}^N \frac{2n_i}{M_1 - 1}}$$

where

n_i is the number of ratings for driver i

σ_i is the standard deviation of these ratings for driver i

σ_{M_1} is the standard deviation of the means for the N drivers

The estimates of reliability and the means and standard deviations of mean ratings on each scale are shown in Table 3 for supervisors' and for associates' ratings. In order to ascertain whether the lower reliabilities of the supervisors' ratings were attributable to the smaller number of supervisory ratings, these estimates of reliability were corrected by application of the Spearman-Brown formula to what they would be if the average number of supervisory ratings were equal to the average number of associate ratings. The corrected estimates of reliability, shown in Table 3, are substantially equal to those obtained on the associates' ratings.

Table 3. Means and standard deviations of mean ratings on 11 driver rating scales, and the reliabilities of the 11 scales

Scale	Supervisors' Ratings (181 Ratees)			Rel. corrected for number of raters	Associates' Ratings (190 Ratees)		
	M	S.D.	Rel.		M	S.D.	Rel.
1	4.78	2.21	.85	.94	4.12	1.49	.94
2	5.56	1.93	.81	.93	4.58	1.58	.94
3	5.47	1.82	.79	.92	4.46	1.49	.94
4	5.84	2.04	.83	.94	4.53	1.55	.93
5	5.87	1.98	.85	.94	4.96	1.67	.94
6	5.92	1.92	.83	.94	4.58	1.55	.93
7	6.01	1.93	.84	.94	4.95	1.44	.93
8	6.00	1.45	.80	.92	4.70	1.29	.93
9	5.55	1.70	.83	.94	4.88	1.52	.93
10	5.68	2.03	.85	.94	4.64	1.44	.93
11	5.65	1.92	.83	.93	4.64	1.48	.93

It can be seen from Table 3 that all of the scales have acceptable reliability; and that there is no choice among the scales on this basis.

Correlation of Mean Ratings with Accident Responsibility Index. Although accident data were insufficient for use as a criterion of safety for the individual, it was considered desirable to select those scales which have highest common variance with an index of accident responsibility, if such an index were obtainable on a sizable portion of the population used. It was found that 28% of the 190 drivers on whom sufficient data were at hand (ratings, estimates of mileage, etc.) had accidents on record. An Accident Responsibility Index was computed on the basis of the accident records of each driver by the following formula:

$$A.R.I. = 1000 \frac{\sum A_i R_i}{M}$$

where A_i = accidents sustained while in present assignment

R_i = responsibility for A_i as estimated from the records on a 5-pt. scale

(1 = no resp., 5 = totally resp.)

M = estimated number of miles driven while in present assignment

^{3/} These estimates were made by the ratees' Commanding Officers, Personnel Officers, or members of their staffs. Persons making these estimates were not included among those asked to make ratings on the scales

The distribution of this index is shown in Table 4.

Table 4. Distribution of Accident Responsibility Index

ARI	f
400-439	2
360-399	0
320-359	1
280-319	1
240-279	1
200-239	0
160-199	1
120-159	1
80-119	4
40-79	13
0-39	29
00	137
	190

Table 5. Correlations between mean ratings and Accident Responsibility Index

Scale	Super. N = 183 Ratees	Assoc. N = 190 Ratees
1	.27	.18
2	.31	.06
3	.23	.14
4	.21	.10
5	.24	.10
6	.20	.10
7	.19	.08
8	.15	.04
9	.23	.08
10	.10	-.02
11	.16	.08

Table 5 shows the correlations between mean ratings and this Accident Responsibility Index. These r 's were expected to be low because of the undoubtedly low reliability of this index. However, this afforded a useful comparison of the scales. It will be noted that the supervisors' ratings have consistently higher correlations with this index than do the associates' ratings. Moreover, the latter show essentially zeros on all the scales, except, perhaps, on scales No. 1 and No. 3. It is also important to note that the lowest correlations between supervisors' ratings and this index are on scales No. 8, 10, and 11. The degree to which any of these correlations might have been affected by raters' knowledge of rateses' accidents cannot be estimated.

Intercorrelations among the Scales and Factor Analysis of the Intercorrelation Matrices. The final guide for the selection of the scales for the final criterion instrument was a factor analysis of the intercorrelation matrices of the mean ratings shown in Tables 6a and 6b. The Thurstone Centroid method was used and rotation effected to the best-fitting orthogonal solution. Table 7 shows the loadings on the three orthogonal factors for each matrix.

Table 6a. Intercorrelations among mean ratings by supervisors on 11 driver rating scales (N = 181 rateses)

Scales	1	2	3	4	5	6	7	8	9	10
1										
2	.78									
3	.65	.73								
4	.78	.73	.70							
5	.84	.79	.71	.86						
6	.81	.73	.63	.85	.85					
7	.73	.62	.60	.77	.80	.86				
8	.45	.44	.45	.46	.47	.50	.46			
9	.71	.78	.67	.76	.76	.69	.59	.43		
10	.46	.49	.46	.47	.50	.49	.44	.63	.51	
11	.65	.53	.44	.62	.67	.67	.67	.52	.60	.66

This apparent discrepancy is considered below under Selection of Raters for final Criterion ratings.

Table 6a. Intercorrelations among mean ratings by associates on 11 driver rating scales (N = 189 raters).

Scales	1	2	3	4	5	6	7	8	9	10	11
1											
2	.81										
3	.81	.82									
4	.81	.84	.78								
5	.84	.83	.79	.86							
6	.84	.79	.75	.88	.90						
7	.79	.77	.75	.85	.89	.90					
8	.44	.54	.61	.57	.57	.58	.57				
9	.76	.83	.78	.83	.86	.82	.84	.65			
10	.56	.70	.61	.66	.71	.67	.67	.63	.74		
11	.69	.72	.67	.75	.80	.80	.79	.61	.79	.77	

Table 7. Factor loadings on orthogonal axes derived from two intercorrelation matrices of 11 driver rating scales

Supervisors' Matrix					Associates' Matrix				
Scale	Factor				Scale	Factor			
	I	II	III	h^2		I	II	III	h^2
1	.80	.30	.03	.80	1	.90	.20	-.07	.87
2	.81	.04	.13	.77	2	.82	.23	.21	.84
3	.76	.01	.15	.70	3	.85	.14	.22	.84
4	.83	.16	.01	.85	4	.78	.44	.04	.87
5	.86	.26	.02	.89	5	.77	.51	.02	.90
6	.80	.44	-.05	.90	6	.74	.57	-.04	.92
7	.68	.54	-.01	.82	7	.71	.57	.02	.88
8	.26	.42	.50	.60	8	.37	.60	.25	.62
9	.74	.09	.24	.74	9	.68	.51	.22	.85
10	.22	.51	.53	.69	10	.46	.45	.51	.71
11	.55	.16	.44	.64	11	.62	.33	.46	.80

Factor I appears to be similar in both matrices. Since it has highest loadings on the scales that deal with overt driving behavior and low loadings on the "non-driving performance" scales (No. 8 and No. 10), we might designate this factor "rated general driving performance." At any rate this is the factor which represents the variance of most of the scales and has low loadings on the suppressor scales (No. 8 and No. 10).

Factor III, is also similar in both matrices. In the associates' matrix it has loadings on scale 10 (Appearance and Military Bearing) and scale 11 (Maintenance). In the supervisors' matrix it also has a loading of .50 on scale No. 8 (Like Dislike). This seems to be an appearance factor, of both the driver and his vehicle, and it would also seem that supervisors like a driver who presents a good appearance.

Factor II is somewhat more difficult to interpret. Moreover, it does not have high loadings on quite the same scales in both matrices. It might represent some aspect of interpersonal relationships or attitude that affects the ratings but is not correlated with overt driving behavior, at least not in the case of the supervisors.

Scales were to be selected which have high loadings on the first factor and low loadings on the other two factors in both matrices.

SELECTION OF THE FINAL SCALES

First, the 6 scales with highest loadings on Factor I were selected. From among these 6, 4 were chosen on the basis of high correlation with the Accident Responsibility Index and low loadings on the other two factors. The scales selected were scale No. 1 (Near Accidents), No. 2 (Reaction to Sudden Changes), No. 3 (Effect of Temper on Driving), and No. 4 (Knowledge of Own Limitations). The distribution of mean ratings on these scales are shown in Figure 1 for associates, and Figure 2 for supervisors.

Scale No. 10 (Appearance and Military Bearing) was selected to precede these four scales in the final booklet. It is not scored but is included to draw off the personal feelings of the rater toward the rates. This scale has low loadings on Factor I and high loadings on Factor III in both matrices. Also, it is less obvious than scale No. 8 (Like-Dislike) which was also designed to be used as a suppressor scale.

DEVELOPMENT OF DRIVING HABIT CHECK LIST

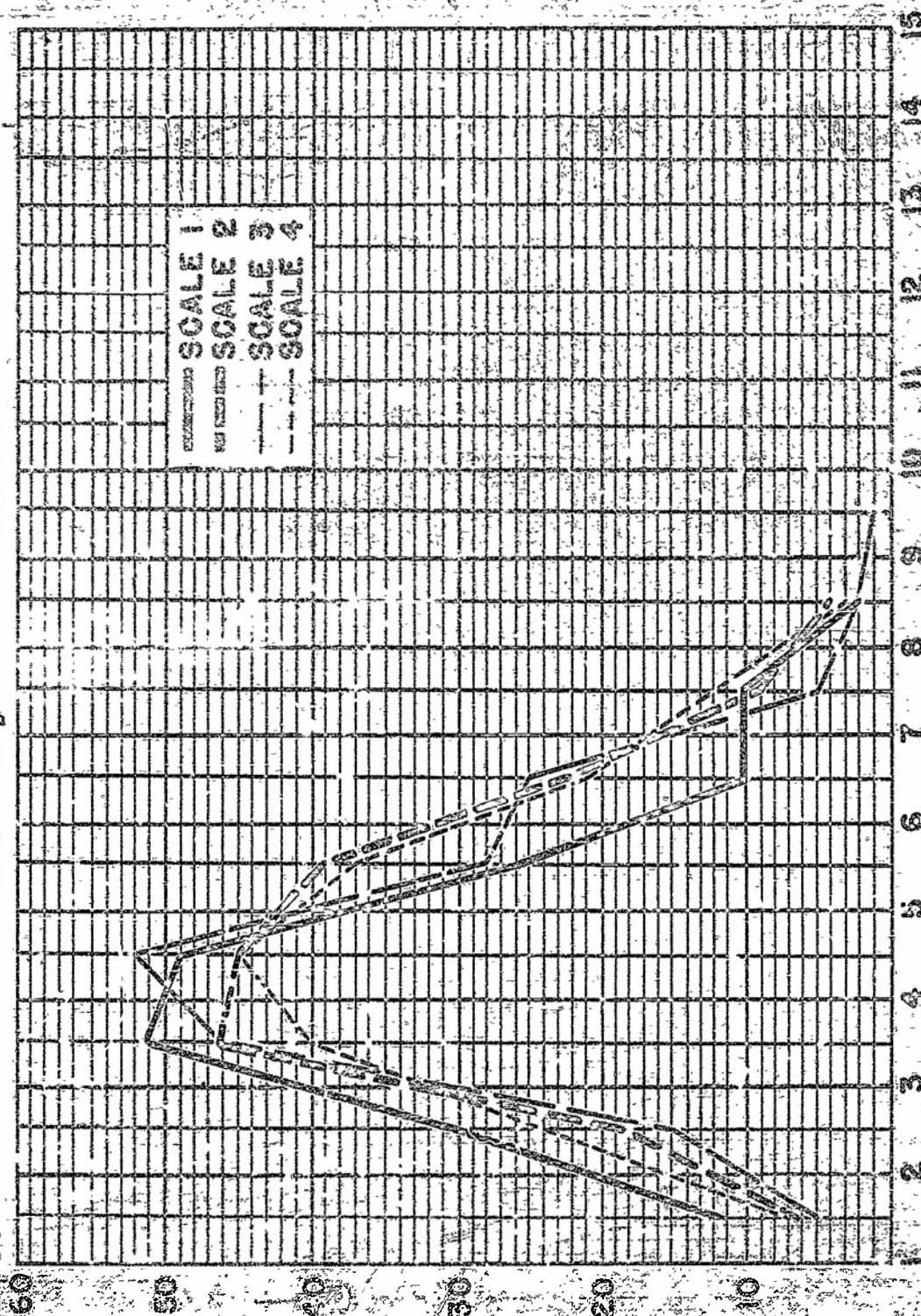
CONSTRUCTION OF EXPERIMENTAL ITEMS

The driving habit check list was developed concurrently with the scales. Suggestions for driving habits which might be considered to be associated with accident causation were sought in the pertinent literature and in consultation with safety personnel, both in the Army and in civilian life. One hundred and five statements of such driving habits were devised. These statements were reviewed with the Army Safety Engineer, the Director of the Pentagon Motor Pool, and several Army drivers in order to assure their clarity, specificity, and the appropriateness of their language to Army drivers.

Distributions Of Mean Ratings Of 189 Drivers

Rated By Associates

FREQUENCY

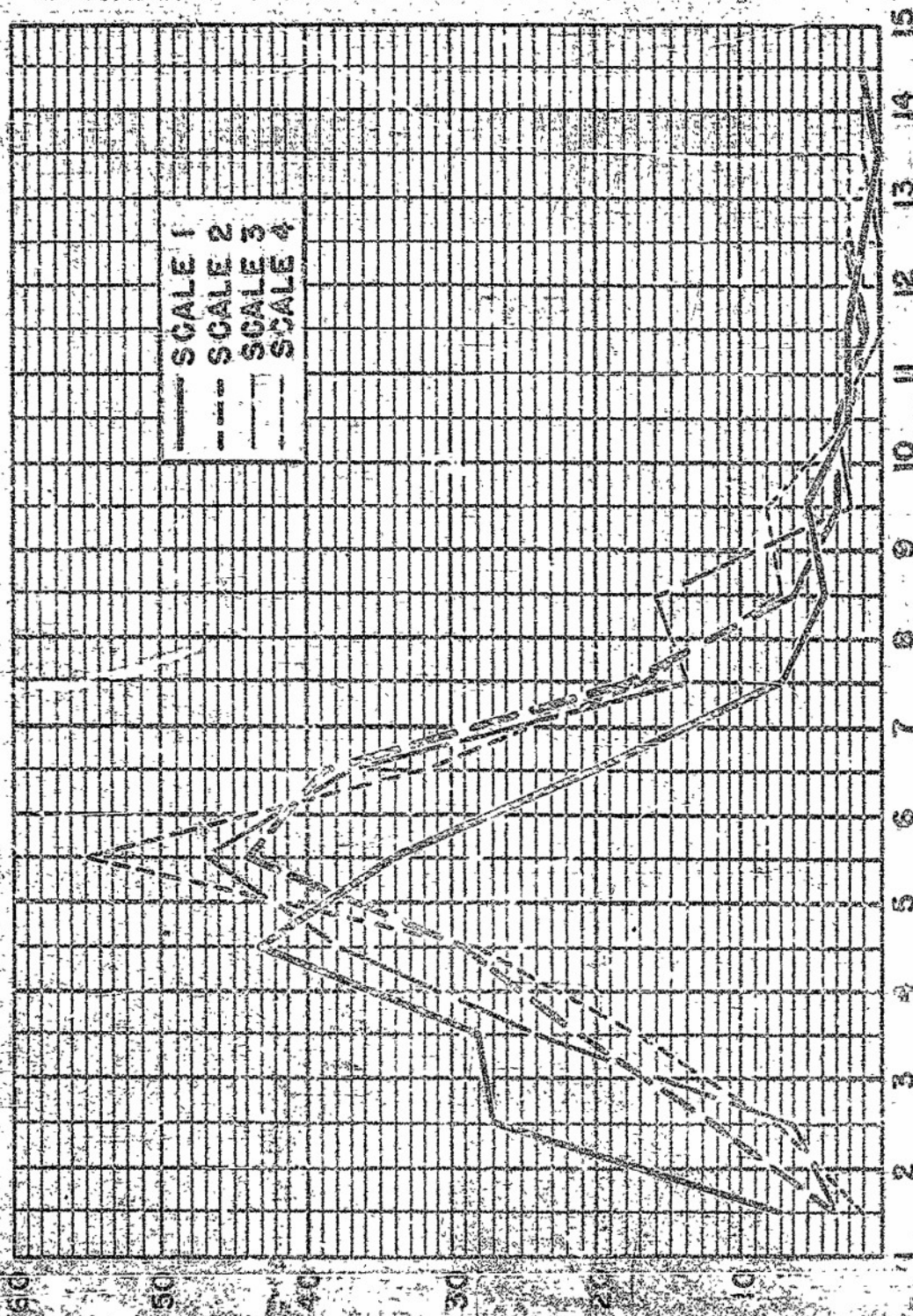


MEAN RATINGS

Figure 1

Distributions Of Mean Ratings Of 181 Drivers Rated By Supervisors

FREQUENCY



MEAN RATINGS

Figure 2

SELECTION OF THE FINAL ITEMS

For an item to serve its purpose in the criterion instrument it would have to be (1) ratable (that is, a statement of observable behavior on which supervisors and associates could rate a driver) and (2) important to safe driving. A study was conducted, therefore, to identify those experimental items considered most important and most ratable by both drivers and supervisors in the Army situation.

The 105 statements were printed in booklet form (PRT 2056) in two parts. Part I, designed to measure ratability (observability), required the subject (consultant) to check one of the following responses for each statement:

- I know I can rate on this.
- I would rate on this but I would not be too sure about it.
- I could not rate on this.

Part II, designed to measure the importance of each habit to safe driving, required the subject (consultant) to check one of the following responses for each statement:

- Very important
- Important
- Not important

This booklet was administered to the same drivers and supervisors who participated in the rating sessions described earlier in this report (158 drivers and 35 supervisors).

On 21 of the items the responses "I know I can rate on this" and "Very important" were checked by at least 49% of the drivers and at least 49% of the supervisors. These items and the percentages of judgments are shown in Table 8. In order to select the 15 items for the criterion instrument from these 21, 6 of the items (Nos. 9, 25, 36, 48, 75, and 84) were eliminated on the basis of the judgments of the investigators.

RELATION BETWEEN SELECTED CHECK LIST ITEMS AND EXPERIMENTAL RATING SCALES

In order to obtain some insight into the relationship between the continua of the rating scales and these 15 items, 8 staff members of the Personnel Research Section were asked to classify the 15 items, following these instructions: "Indicate under which of the rating scales you would subsume each of the driving habits if you were doing the rating." No item was to be assigned by the same PRS judge to more than three different scales. A shortcoming of this method is that primary, secondary, and tertiary assignments receive the same weight. However, the results, shown in Table 9, indicate such a high predominance of assignments to scales 5 and 6 that the findings are readily interpretable.

Table 8. Twenty-one behavior statements rated "very important" and "I know I can rate on this" by 49% or more of drivers and supervisors

Item No	Behavior Statement	Percent Checking			
		"Very Important"		"I Know I Can Rate on This"	
		A*	S*	A*	S*
1.	Breaks the speed limit	39	66	52	60
2.	Drives too fast for road conditions	66	69	51	69
3.	Doesn't stay on his side of the road	73	91	54	49
4.	Ignores stop lights or signs	77	89	61	63
5.	Doesn't give the right of way to other drivers	65	60	58	60
6.	Passes on curves and hills	76	83	49	51
7.	Doesn't signal for stops or turns in advance	57	71	60	63
8.	Doesn't check brakes before driving	71	77	51	51
9.	Doesn't slow down at intersections when he has the right of way	52	63	57	60
14.	Follows other vehicles too closely	59	77	60	69
20.	Pulls away from the curb without looking back for oncoming traffic	66	80	51	51
23.	Takes chances when driving	50	71	51	54
25.	Gets into accidents with other vehicles	64	63	50	74
36.	"Horses around" when he's driving	62	80	50	54
37.	Shows off when driving	61	74	49	60
44.	Doesn't cut wheels to curb when parking on a hill	66	80	49	49
45.	Backs up without looking behind	72	80	53	57
48.	Drives with dirty windshield	55	57	58	71
75.	Swings too wide on turns	55	60	50	54
84.	Drives faster than the other traffic	54	49	51	51
92.	Fails to turn in his vehicle for repairs promptly	60	60	50	77

A* = Associates
S* = Supervisors

Table 9. Assignments of selected check list items to rating scale continua

Scale Number	Item Number															
	1	2	3	4	5	6	7	8	14	20	23	37	44	45	92	
1					1		1	8	1	1	2	1				
2				1					1		1					
3					2											
4				1							1					
5	7	6	8	3	5	8	7	5	6	8	6	6	5	8	1	
6	8	6	3	8	5	5	6	3	5	5	7	7	5	5	3	
7	1	1	1	3	1	1	2			2		3				
8					1											
9		2	2						5				2	1		
10																
11			1					5					2		8	

FINAL COMPOSITE CRITERION

THE FINAL CRITERION INSTRUMENT

The final criterion instrument (FRI 2408) consists of the following:

- (1) The practice rating scale
- (2) The halo scales (not to be scored)
- (3) The four criterion scales
- (4) The driving-habit check list

SCORING OF THE FINAL CRITERION INSTRUMENT

A driver's score on the scales is the mean of ratings received on the four criterion scales, supervisors' and associates' ratings taken together (the basis for this decision is furnished below, under "Selection of the Raters"). His score on the check list is the mean number of checks received. On the basis of the judgments of the investigators with respect to the relative variance contributed by each of these two measures, a weight of 2 for the mean rating and a weight of 1 for the mean check list score was considered to yield about the optimum composite criterion score.

SELECTION OF RATERS FOR FINAL CRITERION RATINGS

Tied in with the problem of selection of scales was the problem of selection of raters. It is evident from the larger means and sigmas of the mean ratings (presented in Table 3 and shown graphically in Figures 1 and 2) and from the consistently higher correlations with the Accident Responsibility Index (Table 5), that the

supervisors' ratings are superior for our purpose. But, since it is often impossible to obtain ratings from more than 2 or 3 supervisors and this criterion is postulated upon having several ratings on each driver, the possibility of supplementing these with the ratings of selected associates was explored.

The first attempt to select among associate raters was made on the basis of grade. The correlations of different grades of associates' ratings with supervisors' ratings are shown in Table 10.

Table 10. Correlations between ratings by supervisors and ratings by different grades of associates

Scale	Ratings by Sgts.	Ratings by Cpls.	Ratings by Pfc's and Priv.	Ratings by all Associates
1	.42	.57	.34	.37
2	.42	.36	.53	.35
3	.40	.51	.38	.42
4	.35	.48	.45	.43
5	.52	.52	.42	.47
6	.53	.49	.41	.46
7	.51	.53	.32	.43
8	.18	.25	.15	.18
9	.42	.45	.37	.41
10	.59	.44	.38	.44
11	.41	.44	.32	.37

These correlations furnished no basis for selection among associate raters.

Findings of another study by the Personnel Research Section(7) indicated that Aptitude Area I score level of raters is more highly related to validity of ratings than is Army grade. Since the raters used in the present study were identifiable only by grade and group, two groups of corporal raters were selected for comparison, Group A being composed mostly of corporals with AA I scores below 90 and Group B being composed mostly of corporals with AA I scores above 90.5. The Pearsonian correlations between Accident Responsibility Index and the ratings accomplished by these two groups of corporals are shown in Table 11. Although correlation on any

⁷The exact mean scores were unobtainable. Group A actually consisted of 33 of 40 corporals whose AA I distribution was known, and Group B consisted of 17 of 24 corporals whose AA I distribution was known. In making the comparisons reported here, the assumption was made that the AA I scores of those--7 in each group--who did not participate in the rating sessions were more or less randomly distributed.

single scale may be inconclusive, the pattern of correlations favors the higher level group with marked consistency. On this basis, the decision was made to obtain criterion ratings from supervisors and only those associates whose AA I scores were at least 90.

Table 11. Correlations of mean ratings with Accident Responsibility Index: Group A with mean AA I below 90, Group B with mean AA I above 90

Scale	Group A*	Group B*
1	.24	.38
2	.15	.28
3	.22	.32
4	.22	.31
5	.15	.30
6	.18	.38
7	.08	.27
8	.11	.17
9	.15	.35
10	.14	.15
11	.15	.27

*In Group A, 33 corporals rated 78 drivers

In Group B, 17 corporals rated 34 drivers

PREPARATION OF REPORT: 7 August 1951

REFERENCES

1. BRANSTFORD, T. Relation of performance on driver's tests to automobile accidents and violations of traffic regulations in the District of Columbia. Unpublished Ph. D. dissertation, Amer. Univ., 1939.
2. BROWN, C. W. and GHISELLI, E. E. Accident proneness among street car motormen and motor coach operators. J. Appl. Psychol., 1948, 32(1).
3. FARMER, E. and CHAMBERS, E. G. A study of accident proneness among motor drivers. Report No. 84, Medical Research Council, Industrial Health Research Board, London: His Majesty's Stationery Officer, 1939.
4. HORST, P. A generalized expression for the reliability of measures. Psychometrika, 1949, 14(1), 21-31.
5. SLOCOMBE, C. S. Consistency of operating efficiency. Person, J., 1930, 8, 413-414.
6. STAFF, PERSONNEL RESEARCH SECTION, The Adjutant General's Office. Personnel research in the Army, VI. The selection of truck drivers. Psychol. Bull., 1943, 40(7).
7. PRS Report 933. A study of ratings obtained from raters with Aptitude Area I scores below 90. 3 April 1952. Personnel Research Section. The Adjutant General's Office, Department of the Army.
8. The accident prone driver. House Document No. 462, Part 6. Bureau of Public Roads. Washington: Government Printing Office, 1938.