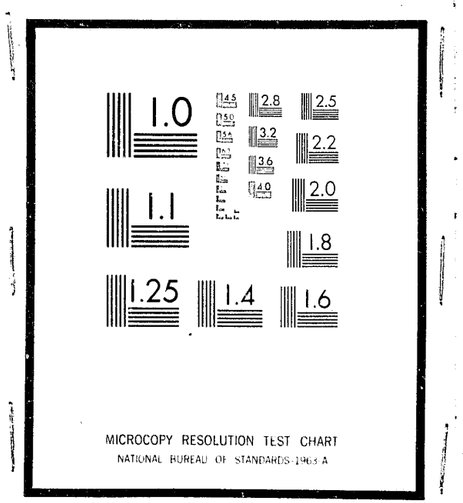

An Evaluation of Polygraph Techniques Currently Practiced by Law Enforcement and Private Polygraph Examiners

Report No. 76-2
Contract 75-NI-99-0001
National Institute of Law Enforcement and Criminal Justice
Law Enforcement Assistance Administration
U.S. Department of Justice

David C. Raskin, Ph.D.
Project Director

Gordon H. Barland, Ph.D.
Project Co-Director

Department of Psychology
University of Utah
Salt Lake City, Utah 84112
April 15, 1976

In recent years there has been increasing interest in the problem of assessing accuracy and reliability of decisions made by field polygraph examiners (Barland & Raskin, 1975; Horvath, 1974; Horvath & Reid, 1971; Hunter & Ash, 1973; Slowik & Buckley, 1975). With the exception of Horvath (1974) those studies have generally found a high rate of agreement between the decisions of the original examiners and independent evaluation of the charts by examiners who had no contact with the subjects and were not given case information which would allow them to make inferences about possible guilt or innocence.

Certain problems arise in the interpretation of those results. First, except for the laboratory study by Barland and Raskin (1975) and half of the cases in the Horvath (1974) study, the sampling of cases has been highly selective so as to include only those cases in which the decision of the examiner had been verified as correct, typically by confession of the guilty person. Thus, the generalizability of those results is open to question. Furthermore, only cases in which the examiner had made a definite conclusion were included in all but the Barland and Raskin (1975) study. However, in the latter study college students served as subjects which limits its generalizability, and the high degree of agreement found in the other studies may have been partially due to the relatively greater ease with which the polygraph recordings could be interpreted. The selection of only cases which had produced conclusive decisions and were later verified is not representative of all cases.

A second problem concerns the procedures employed in the evaluations of the polygraph charts. In all instances, the independent evaluations were performed by examiners who had been primarily trained in the same

technique as that utilized by the original examiner, and the examiners were typically employed by the company or agency which had conducted the original examination. As a result, it might be expected that the uniformity of results was due in part to similar approaches to chart interpretation and particular emphasis on specific aspects of the charts which may be peculiar to the approach of that company or agency.

Recent scientific studies (Barland & Raskin, 1975, 1976; Raskin, 1975) have demonstrated the usefulness of a numerical scoring system for chart interpretation which permits quantification of the outcomes and an assessment of the relative usefulness of the various physiological measures employed. The use of the numerical scoring system in the independent evaluation of polygraph charts randomly sampled from the archives of field polygraph examiners would make it possible to answer questions which have been raised by field examiners concerning the effectiveness of the various measures employed in polygraph examinations (Barland & Raskin, 1973). It also would permit a determination of whether field polygraph examiners base their decisions solely on information provided by the polygraph charts, or if they are influenced by factors in addition to the physiological recordings.

Finally, it is of interest to determine the extent of agreement between independent evaluation of polygraph examinations and the decisions made by law enforcement as compared to private polygraph examiners and the extent to which such agreement varies as a function of the type of crime investigated, the experience and training of the examiners, and the type of test structure employed.

Method

Source of cases

Arrangements were made to sample polygraph examinations from the archives of three state and local law enforcement agencies and four major, private polygraph firms. In order to provide a representative sample of such organizations, the participating agencies and companies included a large, metropolitan police department (Location A), a small city police department which provides polygraph services to outlying communities (Location B), a state police laboratory which serves a major industrial state (Location C), and four private firms (Locations D, E, F, G) in major cities. All but one of the cooperating organizations utilizes a control-question technique; the one exception (G) employs the relevant-irrelevant technique.

Procedure

Visits were made to each of the above organizations by either Gordon H. Barland or David C. Raskin. When they arrived, they sampled cases from the files beginning with the most recent cases. An attempt was made to obtain approximately 60 cases from each location. Cases were selected so as to include examples from three categories: 1) Crimes against people (homicide, assault), 2) economic crimes (robbery, burglary, theft), and 3) sex and drug offenses. All cases were required to meet the criteria of completion of the polygraph test and written indication of the decision by the original examiner. Only tests run on criminal suspects were included in the sample. Victims, witnesses, civil matters, and internal investigations were excluded. When more than one suspect was tested in the same case, only the first suspect examined was included. Using those procedures,

a total of 419 cases were obtained for analysis. The large majority of evaluations was completed before note was made of the decision arrived at by the original examiner. Although the reviewers attempted to complete all evaluations prior to noting the original decision, there were several instances where the original outcome was inadvertently seen prematurely because of the conspicuous manner in which it was marked on the outside of the file folder.

After selecting the cases, the numerical evaluations were performed on the control-question polygraph charts using the methods described by Raskin (1975). That included comparing the magnitude and type of responses which occurred to control and relevant questions, and assigning to each comparison a number which ranged from -3 to +3 for each physiological measure. Negative numbers indicated stronger reactions to the relevant question, and positive numbers indicated stronger reactions to control questions. The scores were summed for each physiological measure for all control-relevant question pairs across all standard charts obtained. Special techniques such as SKY and Yes-tests (Barland & Raskin, 1973) were not scored. If the total score was +6 or larger, the subject was considered truthful; if it was -6 or lower, the subject was considered deceptive; and scores which fell between ±5 inclusive were considered inconclusive. The charts from the relevant-irrelevant tests could not be scored numerically and were given a subjective evaluation using those three categories.

In evaluating the techniques involved, other characteristics of the examinations were noted and included chart quality, chart markings, adequacy of case information, question structure, length of pre-test

interview (when data were available), the source of referral (law enforcement, attorney, employer), and the name of the examiner who conducted the test. Information was also obtained concerning the training and experience of each examiner. A total of 43 examiners were represented in the sample of cases utilized.

Results[1]

Original Decisions

The frequency and percentage of cases where the original examiner arrived at decisions of deceptive, truthful, and inconclusive are shown in Table 1. It can be seen that the various organizations differed in terms of the percent of cases which fell into the three categories for decisions. The use of the deceptive category ranged between 37% and 83%, the truthful category ranged from 12% to 55%, and inconclusives ranged from 0% to 23%.

Agreements Between Original Decisions and Independent Evaluations

Overall results. Using the numerical scores as the criterion for the outcomes of the independent evaluations of control-question tests and subjective evaluations of relevant-irrelevant tests, the extent of agreement between the original decisions and the independent evaluations is shown in Table 2. When agreement was calculated using the three categories of decisions, there was agreement on 239 of 419 cases, or 57.0%. When only decisions of deceptive or truthful were used, there was agreement on 220 of 258 cases, or 85.0%. Thus, when cases in which both evaluations yielded a definite decision were utilized, the extent of agreement was substantially higher than it was when inconclusives were included. When agreement was calculated separately for original

[1] All statistical tests employed a .05 rejection region.

Table 1

Frequency and Percent of Different Decisions Made

by Police and Private Examiners

	(N)	Deceptive		Truthful		Inconclusive	
		f	%	f	%	f	%
Police							
A	(60)	50	83%	7	12%	3	5%
B	(54)	37	69%	17	31%	0	0%
C	(58)	33	57%	22	38%	3	5%
Combined	(172)	120	70%	46	27%	6	3%
Private							
D	(60)	30	50%	16	27%	14	23%
E	(60)	39	65%	21	35%	0	0%
F	(60)	22	37%	33	55%	5	8%
G	(67)	32	48%	35	52%	0	0%
Combined	(247)	123	50%	105	42%	19	8%
Total	(419)	243	58%	151	36%	25	6%

Table 2

Overall Agreement Between Original Decisions

and Independent Evaluations

		Original Decisions			
		Deceptive	Truthful	Inconclusive	Total
Independent Evaluation	Deceptive	180	34	6	220
	Truthful	4	40	0	44
	Inconclusive	59	77	19	155
	Total	243	151	25	419

decisions which were deceptive or truthful, the independent decisions agreed with 97.8% of original deceptive decisions and with 54.1% of original truthful decisions. When inconclusive numerical evaluations were included, the rate of agreement was 74.1% for original deceptive decisions and 26.5% for original truthful decisions. The differences in rate of agreement for deceptive and truthful original decisions was significant when inconclusives were excluded (Z = 8.44) and also when they were included (Z = 7.86). When the rates of agreement for deceptive and truthful decisions excluding inconclusives were each tested against a chance expectation of 50%; agreement on deceptive decisions was significant, [$\chi^2(1) = 168.35$] but agreement on truthful decisions was no better than chance [$\chi^2(1) = .49$].

Since it was not possible to use the numerical scoring system to evaluate the relevant-irrelevant tests obtained from Location G, the rates of agreement were also calculated using only the results obtained for control-question tests. They are presented in Table 3.

The independent decisions based upon numerical evaluations agreed with 98.1% of the original deceptive decisions and with 42.1% of the original truthful decisions. When inconclusive numerical evaluations were included, the agreement was 72.5% with original deceptive decisions and 20.7% with original truthful decisions. The difference between rate of agreement on deceptive and truthful decisions was significant when inconclusives were excluded (Z = 9.66) and when they were included (Z = 8.99). Thus, the numerical evaluations produced significantly higher rates of agreement with deceptive as compared to truthful decisions reached by the original examiner.

Table 3

Agreement Between Original Decisions and Numerical Evaluations of Control-Question Tests

		Original Decisions			
		Deceptive	Truthful	Inconclusive	Total
Independent Numerical Evaluation	Deceptive	153	24	6	183
	Truthful	3	33	0	36
	Inconclusive	55	59	19	133
	Total	211	116	25	352

Comparisons of different locations. In order to assess the extent to which rates of agreement varied across the different organizations in the study, proportions of agreement between independent and original decisions were calculated for each location and are presented in Table 4.

It can be seen that there was a wide range of agreement between independent evaluations and the original decisions made at different locations. The observed differences among locations were significant for decisions which included inconclusives [$\chi^2(6) = 34.57$] and for those which excluded inconclusives [$\chi^2(6) = 15.63$]. The results were combined into Police and Private categories, and the rates of agreement with Police decisions were significantly higher than those for Private examiners when inconclusives were included [$\chi^2(1) = 5.76$] and when inconclusives were excluded [$\chi^2(1) = 7.60$].

Since there was a large variation across locations in the number of instances where the original examiner made a decision and the independent evaluation was inconclusive; the various locations were compared on that basis, and the results are shown in Table 5.

There was a wide range in the proportion of instances where the original examiner had made a decision and the independent evaluation yielded an inconclusive result. Those differences among locations were significant, $\chi^2(6) = 30.20$. However, there was no significant difference between Police and Private examiners, $\chi^2(1) = .67$.

Type of crime. The rate of agreement between decisions based on independent numerical evaluations and those reached by the original examiners are shown in Table 6 according to the type of crime.

Table 4

Agreement Between Independent Evaluations and Original Decisions at Each Location

	Agreements		% Agreement	
	Including Inconclusives	Excluding Inconclusives	Including Inconclusives	Excluding Inconclusives
Police				
A	53/60	50/51	88.3%	98.0%
B	23/54	23/24	42.6%	95.8%
C	34/58	32/39	58.6%	82.1%
Combined	110/172	105/114	64.0%	92.1%
Private				
D	36/60	26/29	60.0%	89.7%
E	26/60	26/34	43.3%	76.5%
F	33/60	29/36	55.0%	80.6%
G	34/67	34/45	50.8%	75.6%
Combined	129/247	115/144	52.2%	79.9%

Table 5

Independent Evaluations of Inconclusive for Original Decisions at Each Location

	N	Original Decisions Evaluated as Inconclusive	% Inconclusive for Original Decisions
Police			
A	60	6	10.0%
B	54	30	55.6%
C	58	16	27.6%
Combined	172	52	30.2%
Private			
D	60	17	28.3%
E	60	26	43.3%
F	60	19	31.7%
G	67	22	32.8%
Combined	247	84	34.0%

Table 6

Extent of Agreement Between Independent Numerical Evaluations and Original Decisions for Three Crime Categories

Type of Crime	Number of Agreements	% Agreement
Against People	46/55	83.6%
Economic	92/106	86.8%
Sex or Drugs	48/52	92.3%

The extent of agreement for the three crime categories was generally high, and the differences among categories were not significant, $\chi^2(2) = 2.52$.

Effectiveness of Physiological Measures

Cases on which numerical evaluation agreed with original decision.

In order to assess the extent to which the three physiological measures of respiration, skin resistance, and cardiovascular activity contributed to the observed results, two types of analyses were performed. The first analysis consisted of selecting those cases on which the decisions based on the independent numerical evaluations were in agreement with the decision of the original examiner at the six locations which utilized control-question tests. They were then separated into deceptive and truthful categories, and the total scores for each component were tabulated. The mean scores for each measure for deceptive and truthful decisions are shown in Table 7.

Analysis of variance indicated that there was a significant difference among the means of the three components for deceptive decisions [F(2/292) = 24.69] but not for truthful decisions [F(2/62) = 1.52]. A Newman-Keuls test indicated that for deceptive decisions the skin resistance measure yielded significantly larger scores than the respiration and cardiovascular measures, which were not significantly different from each other. When the mean for each measure was tested to determine if it was significantly different from zero, all measures for both deceptive and truthful decisions yielded significant results [F>8.57].

The data for each component on agreed deceptive decisions were analysed separately for each location. Truthful outcomes were not analysed since the number of cases at each location was not sufficient to be

Table 7

Mean Numerical Scores for Each Measure in Cases Where the Original Decision and the Independent Numerical Evaluation Agreed

	Respiration	Skin Resistance	Cardiovascular	Total
Deceptive (N = 147)	-3.8	-7.4	-4.7	-15.9
Truthful (N = 32)	+4.5	+4.8	+2.7	+12.0

meaningful. The mean scores for each measure on agreed deceptive decisions at each location are presented in Table 8.

Analyses of variance indicated that skin resistance measures yielded significantly higher scores than respiration and cardiovascular measures at Locations A [$F(2/80) = 26.00$] and E [$F(2/44) = 3.86$], and there were no significant differences among component scores at the other Locations. Separate tests on each component at each Location showed that all measures were significantly larger than zero except for respiration at Location E.

Confirmed cases. A second type of analysis of the effectiveness of components was conducted on all of the cases which were confirmed by confessions or admissions by the person tested or another suspect who confessed to the crime. Since there were only seven cases in which truthful results were confirmed, statistical analyses were performed on only confirmed deceptive cases (N = 56). The mean numerical scores for each measure are presented in Table 9.

Analysis of variance indicated that there was a significant difference among the mean scores for the three measures, $F(2/110) = 8.73$. A Newman-Keuls test showed that skin resistance yielded significantly larger scores than respiration and cardiovascular measures, which were not significantly different from each other. When the mean for each measure was tested to determine if it was significantly different from zero, all three measures yielded significant results, $F(1/110) > 23.34$.

Quality of Examinations

In the process evaluating examinations at each location, certain aspects of test construction, procedures, and quality of polygraph recordings were noted. Many characteristics were noted, including examiner experience,

Table 8

Mean Numerical Scores for Each Measure on Agreed Deceptive Decisions at Each Location

	(N)	Respiration	Skin Resistance	Cardiovascular	Total
Police					
A	(41)	-4.0	-11.4	-6.4	-21.8
B	(20)	-3.2	-5.2	-3.2	-11.6
C	(23)	-4.1	-6.0	-5.6	-15.7
Private					
D	(22)	-3.8	-4.5	-3.6	-11.9
E	(23)	-2.1	-6.3	-3.1	-11.5
F	(18)	-5.8	-7.6	-4.8	-18.2

Table 9

Mean Numerical Scores for Each Measure for Deceptive Results Confirmed by Confessions or Admissions

Respiration	Skin Resistance	Cardiovascular	Total
-2.9	-5.8	-2.8	-11.5

length of pretest interview, review of questions with the subject prior to the actual polygraph test, adequacy of case information and test questions employed, quality of physiological recordings, number of charts obtained, quality and adequacy of chart markings, and use of numerical scoring of charts.

We had planned to analyze the outcomes in terms of extent of agreement as a function of examiner experience. However, only 3 of the 43 examiners had less than one year of experience, and they accounted for only 15 of the 419 cases sampled. Therefore, the number of cases obtained from inexperienced examiners was insufficient to justify such an analysis.

Of the seven locations visited, all but Location G utilized control-question technique. That is also the only location where the questions on the examination were not reviewed with the subjects prior to the polygraph test. The lack of control questions also made it impossible to numerically evaluate the polygraph charts from that location, and the results from Location G could not be included in many of the analyses.

There were a number of characteristics which varied widely among the locations visited. Each location is described below, making note of the findings with regard to those characteristics.

Location A. The overall quality of examinations was clearly the highest among those sampled in this study. The examinations sampled were conducted by eight different examiners who were trained at the U.S. Army Polygraph School (3), Gormac (4), and in-service training with Gormac Certificate (1). In the last few years a great deal of in-service training has been conducted among the staff. All of the examiners had more than one year of polygraph experience. Case

information was routinely obtained by the examiner from an interview with the investigating officer prior to the examination, and a pretest interview averaging 74 minutes was conducted with each subject and included a complete review of all questions to be asked on the test. The test structure was either a Reid or federal modification of the Backster Zone Comparison control-question test, and the control questions in all tests were clearly separated from the issue tested by using time exclusions. All tests consisted of a minimum of three charts and employed a number test between the first and second charts. Some tests included a silent answer test and/or a yes test. The results of the yes tests were not included in our independent evaluations. The question structure was uniformly excellent, and the quality of polygraph charts and chart markings were also excellent. There were a number of instances where the skin resistance sensitivity was too high. Very extensive and complete records and files were kept on all cases, but numerical evaluation of charts was almost never performed. All examinations were recorded in their entirety on audio tape.

Location B. The overall quality of examinations was generally high. Both polygraph examiners at this location were trained at the National Training Center of Lie Detection and had more than one year of experience in conducting examinations. The total number of examinations conducted at this location was very small, being less than 300. The primary reason for the large number of inconclusive results based on numerical scores was that only two charts were usually obtained by the original examiner. The quality of the charts was generally high, with marks indicating both question onset and offset, adequate spacing between questions, and good

amplitude for each of the physiological measures. The test structure used was the Arther control question test. A guilt complex type question, the Arther known truth question, was employed with most tests. All tests consisted of a minimum of two charts. The question structure was generally excellent. Complete records and files were kept on all cases, but numerical evaluation was never performed.

Location C. The overall quality of examinations was good. The examiners at this location were trained at John Reid & Associates. The Reid control question test was used for all examinations. Difficulties encountered in interpreting the charts at this location included the fact that neither question onset nor question offset was marked on the charts. The only way of identifying the question locations was from the question number written on the chart and the point at which the subject answered the question. The case information usually consisted of police reports and an interview with the investigators. All tests consisted of a minimum of 2 charts; usually 3 or 4 were obtained. A number test was usually given, generally between the 1st and 2nd charts. A mixed question test was usually given, and a silent answer test and yes test were occasionally included toward the end of the testing. The physiological recordings were generally of good quality, but the skin resistance amplifier was sometimes set too low, usually by one examiner. Complete records and files were kept on all cases, but numerical evaluation was never performed.

Location D. The overall quality of the exmainations was good. A total of four different examiners all trained at the Backster School conducted the examinations, and one of them had less than one year of

polygraph experience. Case information was usually obtained from a telephone or in-person conference with the attorney, employer, or police investigator prior to the examination, and a pretest interview averaging 73 minutes was conducted with each subject and included a complete review of all questions to be asked on the test. The test structure always included at least two charts of a Backster Zone Comparison control-question test and often included a SKY test which was not included in our numerical evaluation. Only 23 of the 60 examinations included three or more standard, control-question charts which greatly contributed to the large number of cases in which the original and/or independent evaluation was inconclusive. No number tests or other stimulation tests were used. The question structure was generally good, although there were several deviations from standard Backster question structure. The quality of polygraph charts was generally good except for many instances where the skin resistance sensitivity was too high. Chart markings were also good except for 31 examinations by one examiner whose failure to mark the onset of questions made the interpretation of the charts difficult. Very few examinations were numerically evaluated by the examiner, and many files did not include a written indication of the decision by the examiner.

Location E. The overall quality of examinations was adequate. The examiners were trained at the Keeler Institute, the National Training Center of Lie Detection, and the American Institute of Polygraph Technology. The Reid Control question test was used for all examinations at this location. The case information usually consisted of verbal information from the requestor, occasionally supplemented with documentation. All tests consisted of a minimum of 2 charts; often 3 or

4 were obtained. A number test was usually given, generally between the 1st and 2nd charts. A yes test, yes-no test, or a silent answer test was usually included, as was a mixed question test. The physiological recordings were of generally good quality, but the skin resistance amplifier was often set too high or too low. Seven of the eight disagreements between the original examiner and the reviewer occurred with one examiner whose charts showed considerable improvement the longer he was employed at this location, but who was terminated because of the poor quality of his examinations. Complete records and files were kept on all cases, but numerical evaluation was never performed in spite of the fact that all examiners had been trained in it.

Location F. The overall quality of examinations was very good. The examiners at this location were trained at John Reid & Associates. The Reid control question test was used for all examinations. Difficulties encountered in interpreting the charts at this location included the fact that neither question onset nor question offset was marked on the chart. The only way of identifying the question locations was from the question number written on the chart and the point at which the subject answered the question. The skin resistance amplifier was often turned too low or was turned off completely on some of the charts. All tests consisted of a minimum of two charts plus a number test, usually given after the first chart. Usually, 3 or 4 charts were obtained. A mixed question test, a yes test, or a silent answer test was usually administered. Complete records and files were kept on all cases, but numerical evaluation was never performed.

Location G. The relevant-irrelevant test was used for all examinations at this location. The examinations were conducted by only one examiner trained in the relevant-irrelevant technique. The case information usually consisted of verbal information from the requestor. Most tests consisted of two charts. No number test was given, nor were any special purpose tests such as the yes test or silent answer test used. The physiological recordings were of generally poor quality with the skin resistance amplifier invariably turned too low and the cardio tracing of such a large amplitude that it often hit the upper pen stop midway through a test. No efforts were made to adjust the pen downward.

Discussion

The 85% agreement between the decisions made by the original examiner and those made by the independent evaluators was quite high. However, when inconclusive determinations were included, agreement dropped to only 57%. That was due to the fact that the original examiners used the inconclusive category on only 6% of the cases, whereas the independent evaluators found 37% of the tests to be inconclusive. Further analyses indicated that the independent evaluations produced higher agreement on cases where the original examiner had judged the subjects to be deceptive as compared to original decisions of truthful. Although deceptive decisions based on independent numerical evaluations agreed with examiner decisions of deceptive 98% of the time, agreement on truthful decisions did not exceed chance levels.

The overall results seem to indicate that many examiners failed to use the inconclusive category when it was appropriate to do so, and they often appeared to be concluding that subjects were truthful when the

polygraph charts did not provide a scientifically-supportable basis for that decision. Obviously, some examiners frequently rendered decisions on the basis of information not contained in the physiological recordings. It must be presumed that in those cases they were basing their decisions on behavioral cues, case information, previous experience, intuition, or other factors.

There was a very large variation across the seven locations in the extent to which their original decisions agreed with the independent evaluations. In general, significantly greater agreement was found with police as compared to private examiners. When inconclusives were excluded, agreement ranged from 98% agreement with Police Location A to 76% agreement with Private Location G. It also was the case that Police Location A had the lowest rate of examinations where the original examiner arrived at a decision which was considered inconclusive by the independent evaluator. Those rates ranged from 10% at Location A to 56% at Location B. A great deal of those instances were due to an inadequate number of charts.

The analyses of agreement on outcomes as a function of type of crime involved yielded no significant differences among the categories of crimes against people, economic crimes, and sex and drug crimes. It appears that equally conclusive polygraph results are obtainable on a wide variety of crimes.

The comparisons of results obtained with the three standard physiological measures yielded data similar to those reported from scientific studies (Barland & Raskin, 1975, 1976; Podlesny & Raskin, 1976; Raskin,

1975). The results indicated that measures of skin resistance, respiration, and cardiovascular activity contributed significantly to the outcomes at all locations where numerical evaluations could be performed, except for respiration at one location. Evidence was also obtained which indicates that the skin resistance response is clearly the most effective measure with deceptive subjects and is as effective as respiration and cardiovascular measures with truthful subjects. Those data are in conflict with the reports by some field examiners that the skin resistance response is not effective in field polygraph situations (see Barland & Raskin, 1973). Such opinions have been strongly expressed by a number of examiners at the locations sampled, including one location where skin resistance was found to be significantly more effective than the other two measures.

The overall results of the study indicate relatively good performance by field polygraph examiners, especially those at the law enforcement agencies included in the study. However, there were a number of findings which clearly point to a need for improvement in the following areas:

1. All examinations on criminal suspects should obtain at least three charts when control-question tests are employed.
2. The examiner should be careful to adjust the sensitivity of the skin resistance measure so as to obtain clearly interpretable recordings, and he should mark the onset of questions as well as the end of questions and the point of answer.
3. Audio tape recordings would be helpful in providing a permanent record of the examination procedures. Such a record would be useful in resolving questions which may arise later and in providing a check on the accuracy of markings on the polygraph charts and the actual

wording and timing of questions asked.

4. All standard, control-question charts should be numerically evaluated using a system similar to that supported by scientific studies. The rate of disagreements would be reduced and the accuracy rate might be increased.
5. Examiners should be willing to use the inconclusive category more often than they presently use it. Too frequently they render decisions which are not justified by the polygraph charts and may be the result of undue reliance on questionable sources of information. The use of behavioral assessments of the subject should be approached with caution and then should be used only as protection against possible errors. Behavioral cues are not presently justifiable as the basis for definite conclusions about truth and deception.
6. Until there is scientific evidence which indicates that the relevant-irrlevant technique has overcome a number of psychophysiological objections which have been raised against it (Podlesny & Raskin, 1976), the routine use of relevant-irrelevant tests with criminal suspects should be viewed with great caution and should be replaced by scientifically supportable techniques such as control-question or guilty-knowledge tests.

References

Barland, G. H., & Raskin, D. C. Detection of deception. In W. F. Prokasy & D. C. Raskin (Eds.), Electrodermal activity in psychological research. New York: Academic Press, 1973.

Barland, G. H., & Raskin, D. C. An evaluation of field techniques in detection of deception. Psychophysiology, 1975, 12, 321-330.

Barland, G. H., & Raskin, D. C. Validity and reliability of polygraph examinations of criminal suspects. (Report No. 76-1, Contract 75-NI-99-0001, U.S. Department of Justice). Salt Lake City, Utah: University of Utah, Department of Psychology, March, 1976.

Horvath, F. S. The accuracy and reliability of police polygraphic ("lie detector") examiners' judgments of truth and deception: The effect of selected variables. Unpublished doctoral dissertation, Michigan State University, 1974.

Horvath, F. S., & Reid, J. E. The reliability of polygraph examiner diagnosis of truth and deception. The Journal of Criminal Law, Criminology and Police Science, 1971, 62, 276-281.

Hunter, F. L., & Ash, P. The accuracy and consistency of polygraph examiners' diagnoses. Journal of Police Science and Administration, 1973, 1, 370-375.

Podlesny, J. A., & Raskin, D. C. Physiological measures and the detection of deception. Submitted for publication, 1976.

Raskin, D.C. Psychopathy and detection of deception in a prison population. (Report No. 75-1, Contract 75-NI-99-0001, U.S. Department of Justice). Salt Lake City, Utah: University of Utah, Department of Psychology, June, 1975.

Slowik, S. M., & Buckley, J. P. Relative accuracy of polygraph examiner diagnosis of respiration, blood pressure, and GSR recordings. Journal of Police Science and Administration, 1975, 3, 305-309.

END