Journal of Applied Psychology
1982, Vol. 67, No. 2, 131-137

A Method for Evaluating the Use of the Polygraph in a Real-Life Situation

Avital Ginton, Netzer Daie, and Eitan Elaad
Scientific Interrogation Unit, Israel Police Headquarters, Jerusalem, Israel

Gershon Ben-Shakhar
Hebrew University, Jerusalem, Israel

Previous attempts to evaluate the polygraph in a real-life situation suffered from several defects. The present study attempts to eliminate these defects by meeting six necessary conditions for an acceptable real-life validation of the polygraph. The procedure devised guarantees the objective identification of "liars," without jeopardizing the real-life appearance of the experimental situation. Fifteen subjects participated in the experiment; two of them actually cheated on a test. All subjects went through a standard polygraph test using the control questions method. Each subject was evaluated by three polygraphers. One had access to the polygraph charts only, one observed the subject's behavior but not his charts, and a third, who conducted the interrogation, had both kinds of information. The evaluations of all three were compared with the criterion. The evaluations that were based both on behavior observation and on the physiological charts were superior to those based on either type of information alone. However, the evaluations based on the physiological information alone were not superior to those based on the behavioral information alone.

Polygraphic interrogation is a very important area of applied psychology, both in terms of the extent of its use and in terms of its social ramifications (Lykken, 1974). However, very few validity studies concerning the use of the polygraph in real-life situations have been done, and the interpretation of these few studies is highly controversial (Lykken, 1979; Raskin & Podlesny, 1979).

The interrogation method most frequently used by polygraphers today is the Control Question Technique (CQT; Podlesny & Raskin, 1977; Reid & Inbau, 1966). Briefly, the CQT utilizes two types of questions for interrogations: relevant questions, which pertain to the crime under investigation, and control questions, which are specially formulated in a pretest interview so that the suspect is likely to be deceptive or very concerned about them. A typical relevant question in a theft case might be: "Did you steal the diamond ring from the safe of Mr. Smith last Friday?"; a typical control question might be: "Have you stolen something from someone who trusted you?"

Physiological changes are continuously recorded during the interrogation. The responses to the relevant questions are compared with the responses to the control questions. The typical classification rule would be to identify a suspect as deceptive if he or she tends to react more to the relevant questions, and to identify him or her as nondeceptive if he or she tends to react more to the control questions. When there is no clear tendency in either direction the result of the test would be labeled inconclusive.

This article is based on a master's thesis conducted by Netzer Daie under the supervision of Gershon Ben-Shakhar.

We would like to thank Meir Kaplan, head of the Criminal Identification Division of the Israeli Police, for giving permission to carry out this project; Meir Gross, head of the Instruction and Scientific Interrogation Section, for his helpful advice and assistance; Ester Brown and Ilana Elaad for their assistance in testing the subjects; and the Scientific Interrogation Unit staff for its helpful cooperation. We would also like to thank Maya Bar-Hillel for her help, in clarifying the presentation of this paper, and David Lykken and an anonymous reviewer for their helpful comments and careful reviews.

E-mail: ginton@zahav.net.il

The theoretical rationale behind this rule is that under proper pretesting conditions, it is possible to set up a situation in which a suspect who is truthful concerning relevant items would be more concerned about the control questions and would produce greater responses to them than to the relevant questions. On the other hand, deception on relevant questions should make the subject more responsive to relevant than to control questions. (Podlesny and Raskin, 1977).

Lykken (1978, 1979) doubted the soundness of this rationale. In particular, he doubted whether autonomic responses to the control questions provide a reasonable estimate of what the subject's responses to the critical questions would have been if he or she were answering truthfully (Lykken, 1979). The empirical aspect of the Lykken–Raskin dispute relates to the validity of the CQT. Podlesny and Raskin (1977, p. 787) estimated that 88% to 96% of the suspects tested are classified accurately. Lykken (1979), on the other hand, claims that in real-life situations, the polygraph test has an accuracy of 64% to 71% against a chance expectancy of 50%, when the polygraph charts are scored blindly.

The sources for this lack of clarity regarding a strictly empirical question are mainly methodological. The most crucial problem is the lack of a criterion. In real-life situations the truth is seldom available. Three approaches to overcome this difficulty have been suggested:

(1) The use of a mock crime experiment in which subjects simulate commiting a crime (Barland & Raskin, 1975; Dawson, 1980; Podlesny & Raskin, 1978; Raskin & Hare, 1978)—this type of study is inadequate for assessing the validity of the CQT in real criminal investigations mainly because realistic fear of failure plays no role whatever in these studies (Lykken, 1979).

(2) The derivation of criterion from a panel of legal experts who have access to all the information relevant to the case except for the polygraph results (Bersh, 1969; Barland & Raskin, Note 1). There are several problems with this technique as well: (a) One can never be certain that the panel decision is indeed correct. (b) The panel and the polygraph interrogator might not be independent, since both may be exposed to the same prior information. (c) There is often a selection bias in this type of study, since the files cannot be used unless they include sufficient information for the panel to reach a conclusion (Bersh, 1969).

(3) The use of confessions to determine who was guilty and who was innocent in a given crime, thus working backward from the criterion (e.g., Horvarth, 1977; Horvarth & Reid, 1971). This technique introduces a substantial sampling bias, since there might be a relationship between the polygraph results and the probability that a suspect will confess.

Another methodological problem in many polygraph validity studies is that of contamination. Since the polygraph operator working with the CQT usually has access to all prior information about the suspect and the case and must interview the suspects, it is not possible to attribute his or her judgment to the physiological responses alone. Some researchers (Horvarth, 1977; Horvarth & Reid, 1971) tried to overcome this problem by using "blind" polygraphers to evaluate the polygraph charts after the interrogation was completed.

The goal of the present paper is to suggest a new design for validating the CQT that will overcome most of the above mentioned difficulties. To validate a claim for the ability to detect which subjects did or did not commit some act it is desirable that: (a) the act (e.g. deception) be authentic, and freely undertaken rather than simulated; (b) it be independently ascertainable which subjects actually committed the act; (c) the subjects believe that the interrogator does not know who committed the act; (d) the subjects be genuinely concerned about the outcome of the polygraph test; (e) the polygrapher have access only to the polygraph charts; and (f) the polygrapher not know the proportion of guilty and innocent subjects in the sample.

The present article suggests an experimental design that satisfies these conditions. In addition, an effort will be made to estimate the role of behavioral clues in the polygrapher's evaluations. That is, we attempt to separate the pure contribution of the physiological information from the benefits of observing the interrogated subject.

Method

Subjects

Twenty-one males serving in the Israeli police participated in the experiment. Their mean age was 29.2 years. At the time of the experiment, all of them were participating in a police course.

Apparatus

The polygraph examinations were conducted in three standard, sparsely furnished examination rooms. Each room contained a table with a built-in polygraph, two chairs, a carpet on the floor, bare walls without windows, and a one-way mirror through which an observer could observe and listen to the polygraph test from outside.

Three Lafayette 76056 field model polygraphs were used in the experiment. Each polygraph recorded respiration, galvanic skin response (GSR), and cardiovascular activity. Respiration was recorded by two pneumatic tubes positioned around the thoracic area and abdomen. The GSR was recorded from two stainless steel electrodes attached to the index and fourth fingers of the subject's left hand (in one examination room the subject's sitting position demanded the attachment of the GSR electrodes to the subject's right hand). Following standard field practice, no electrode paste was used. Cardiovascular activity was recorded by a pneumatic pressure cuff positioned around the upper portion of the subject's contralateral arm. The cuff was inflated to a pressure of 60 mm Hg.

Procedure

At the first stage of the experiment, paper-and-pencil tasks were administered to the subjects. They were presented as aptitude tests that the subjects had to take as part of their course. In one of the tests the subject had to enter solutions in a 5×5 matrix according to certain written instructions. Beneath the answer sheet there was a hidden chemical page that received an impression of what was written on the answer sheet. After completion of the test, the answer sheet was separated from the rest of the pages and handed back to the subjects. The correct answer keys were then handed out and the subjects were asked to score their own tests. At this stage, subjects could improve their score by adding correct solutions to previously empty cells, or by changing their original wrong answers. It was possible to know whether and how a subject had tampered with his answer sheet by comparing it to the chemical copy that included his original answers. It turned out that seven of the subjects cheated while scoring their tests.

After several days the subjects were told that there were suspicions that some of them cheated when scoring the aptitude test. They were offered an opportunity to take a polygraph test, and it was made clear that their future careers in the police might depend on the outcome of such a test. All 21 subjects initially agreed to take the polygraph test, but at a later stage one "guilty" subject did not show up for the test and two subjects (one guilty and one "innocent") refused to take it. Three other guilty subjects confessed before taking the polygraph test. As a result there were only 15 subjects that went through the polygraph interrogation, of whom only two had actually cheated in their scoring.

The polygraph examinations took place at the police laboratories in Jerusalem. There were seven interrogators, each of whom had at least 1½ years of experience as a professional polygrapher. Of these, three polygraphers were involved with each interrogation: One conducted the interrogation, another observed and heard the pretest interview from an adjacent room through a one-way mirror, and a third later evaluated the polygraph chart.

A typical series of questions used in the present experiment is as follows:

(1) Is your name _____? (the subject's name was inserted)
(2) Do you wear a shirt?
(3) Do you intend to lie on this test?
(4) Is today _____? (the appropriate day was inserted)
(5) Did you copy anything from the correct answer sheet onto your own sheet?
(6) Did you ever deliberately cheat in order to succeed?
(7) Are you now in Jerusalem?
(8) Did you make any change on your own answer sheet using the correct answer sheet?
(9) Have you ever taken advantage of trust placed in you?

Questions 5 and 8 are considered relevant questions; Questions 6 and 9 are the control questions; Questions 1, 2, 4, and 7 are considered irrelevant. Question 3 appears as relevant, but is not treated as such in the analysis; it is considered to be a general stimulating question. The series of questions was repeated at least three times. As in most standard applications of the CQT, a card test procedure was introduced after the first series of questions. This procedure involved the sequential presentation of several cards, one of which had been chosen in advance by the subject; the subject replies "no" to each of the cards when asked if it is the chosen card. (For a detailed description of the card test procedure and its rationale, see Reid & Inbau, 1966, pp. 27–28).

The polygraphers knew the nature of the experiment but were blind as to the guilt or innocence of the subjects, and did not know the proportion of guilty subjects in the group.

After all 15 polygraph tests were completed, the subjects were debriefed. They were given an opportunity to express their attitudes and feelings about the experiment, and the whole issue was discussed. The goal of the research and its importance was explained to the subjects and they were assured that all information concerning individual subjects was to be kept in strict confidence and to be used only for the research analysis. All subjects accepted the experiment and its justification.

Two months after completion of the experiment, the 15 charts were sent out for an additional evaluation by five of the original polygraphers and three additional polygraphers. To assure that there would be no prior knowledge of the distribution of guilty and innocent subjects, 10 additional charts were sent along with the 15 original ones. There were no markings of any kind on

Table 1
Distribution of Global Evaluations and of Evaluations Based on Field Scoring Technique For Guilty and Innocent Subjects

	Global evaluations		Evaluation based on field scoring technique	
Evaluations	Guilty	Innocent	Guilty	Innocent
Original interrogator				
Guilty	2	2	1	1
Innocent	0	11	0	6
Inconclusive	0	0	1	6
Blind chart evaluator				
Guilty	1	3	1	1
Innocent	1	7	0	5
Inconclusive	0	3	1	7
Observer				
Guilty	0	2	—	—
Innocent	2	11	—	—

the charts, and as far as we could ascertain, there was no way to tell the 10 additional charts from the 15 original ones. Each chart was blindly evaluated by each of the eight polygraphers.

Results

The three polygraphers who participated in each interrogation gave an overall global evaluation concerning the guilt or innocence of each subject. These evaluations were based on the charts in the case of the blind chart evaluator, on the behavior of the subject in the case of the observer, and on both the chart and the behavior in the case of the interrogator. The outcome of these evaluations is presented in the left half of Table 1.

Both the original interrogator and the blind chart evaluator evaluated the polygraph charts a second time using the field score technique (Barland and Raskin, Note 1). This technique involved comparisons of the response to each relevant question with the response to the nearest control question, separately for each of the three physiological measures. These comparisons were scored from −3 to 3, depending on the intensity and the direction of the difference, and a final score was derived by summing these partial scores across all measures and all relevant questions. Total scores from −5 to +5 were deemed inconclusive, whereas more extreme scores determined a guilty or innocent identification as appropriate. The result of this more objective evaluation is presented in the right half of Table 1.

The eight polygraphers who received the chart 2 months after the completion of the experiment also evaluated the charts by these two methods. The distribution of their evaluation using the global method is presented in Table 2, and that of the field scoring technique in Table 3.

The evaluations of the 10 additional charts were also analyzed by the two methods. The global method produced an average of 5

Table 2
Distribution of Blind Chart Evaluations (Guilty, Innocent, or Inconclusive) of Eight Polygraphers 2 Months After Completion of Experiment

	Guilty subjects			Innocent subjects		
Polygrapher	Guilty	Innocent	Inconclusive	Guilty	Innocent	Inconclusive
1[a]	2	0	0	1	12	0
2[a]	2	0	0	1	11	1
3[a]	2	0	0	1	11	1
4	2	0	0	2	11	0
5	2	0	0	2	11	0
6	2	0	0	4	9	0
7	1	1	0	3	10	0
8	2	0	0	4	9	0
n	15	1	0	18	84	2
%	94	6	0	17	81	2

[a] Polygraphers 1–3 did not participate in the first part of the experiment.

Table 3
Distribution of Evaluations Based on a Field Scoring Technique Made by Eight Polygraphers

Subjects	Evaluations			
	Guilty	Innocent	Inconclusive	Total
Guilty				
n	11	0	5	16
%	68.8	0	31.2	100
Innocent				
n	8	38	58	104
%	7.7	36.5	55.8	100

guilty, 3 innocent, and 2 inconclusive evaluations. The field scoring technique produced an average of 3.5 guilty, 1.5 innocent, and 5 inconclusive evaluations.

Discussion

The present study attempted to validate polygraphy in a real-life situation; that is, one where the guilty parties commit the "crime" out of free choice, and are seriously and genuinely concerned about the results of the polygraph test. It can be argued that although the study simulated a real-life situation for the subjects, it was not a real-life simulation for the polygraphers, since the polygraphers knew in advance that their judgments would have no consequences for the subjects. Note, however, that the polygraphers' judgments had significant consequences for themselves. Their validity as polygraphers was on the line. Clearly, therefore, they were motivated in this immediate and objective feedback situation to do their best. Furthermore, even if the payoff structure in this artificial (from the polygrapher's perspective) task was different than the one encountered in real life, it could only have differed in a manner that reduces response bias by making payoffs more symmetrical. Thus we believe this study to be a credible test of polygraphy, carried out under adequate methodological control.

Unfortunately, however, the very nature of the experiment makes it difficult to obtain large samples, and even more so to prevent subjects from dropping out or confessing. Hence, the results must be treated with some caution.

From inspection of Table 1 it seems that the accuracy of the blind chart evaluators is not better than the accuracy of the polygraphers who were not exposed to the charts at all and who based their evaluations on their observations of the subjects' behavior only. In both cases, there were four errors out of 15 cases, and indeed in the case of the blind chart evaluators, there were also three inconclusive decisions. We may add that all four errors made by observers of behavior were made by a single observer, whereas all 11 correct identifications were made by the three other observers who participated in the experiment. This may suggest that observation of behavior could turn out to be a valid tool for detecting deception, perhaps supplementing the physiological information. Indeed the interrogators, who had access to both types of information, did better than the evaluators who used either type alone.

At this stage we cannot identify the behavioral clues that formed the basis for the judgments, nor can we account for the differences among the different observers. Further research is needed to clarify this question. In any event, it is clear that any attempt to estimate the validity of the polygraph per se must use a blind procedure, in which the person who scores the charts is denied access to any nonphysiological information.

In contrast to the rather poor performance of the blind chart evaluators immediately after the interrogation, the evaluations made 2 months later seem to be much more accurate. In the initial evaluations, the rate of correct identifications is quite low—1 out of 2 guilty subjects, and 7 out of 13 innocent subjects. This is even lower than the rates reported by Hovarth (1977) and Barland and Raskin (Note 1). In the delayed situation, on the other hand, the rates of correct identifications are reasonably high—81% of the evaluations of charts belonging to innocent subjects and 94% of those of guilty subjects.

This rather surprising difference may be explained by the fact that right after the experiment, the polygraphers worked under time pressure, whereas in the delayed situ-

ation, they analyzed the charts at their leisure. Another possible explanation for this difference could be due to the polygraphers. Polygraphers 1, 2, and 3 in Table 2 did not participate in the original experiment, whereas the other five polygraphers did. Since these three polygraphers achieved slightly better results than the other five, the improvement after 2 months cannot be attributed merely to the effect of the five rerun polygraphers. Neither can the increase in accuracy be explained by the fact that the polygraphers knew the base rates of guilty and innocent subjects in the sample, since they were exposed to 10 additional charts taken from a source unknown to them. Indeed, these 10 were more frequently evaluated as guilty than as innocent.

In any event, the results of the second scoring must be treated with great caution. Although the polygraphers in the original experiment did not receive any explicit feedback beyond being told how many cases they correctly identified, we cannot rule out the possibility that some sort of feedback was implicitly conveyed to them. If this happened while the charts were still fresh in their minds, then it would be natural for them to think about their errors, to try to recall features of the charts that had led them astray, and so on. Such recollection might have aided them in avoiding these same mistakes 2 months later. Furthermore, it is possible that the polygraphers discussed the issues with each other and with the three new polygraphers, though we have no knowledge of such discussion. If so, their judgments may not have been totally independent, although they performed the task individually. These possibilities, if true, might have artificially inflated the polygraphers' accuracy the second time around, but we tried our best to safeguard against them when concluding the experiment.[1]

Although it is more objective, the field scoring technique recommended by Barland and Raskin (Note 1) and used by many others did not improve the accuracy of the polygraphers in the present study. Interviews with some of the polygraphers who participated in the present experiment revealed that whereas the field scoring technique is based only on comparison of responses to control and relevant questions, the global evaluations take into account other features of the chart as well; for example, several polygraphers mentioned that they were influenced by differences in responsivity before and after the card test. An effort should be made to quantify all the relevant aspects of the charts and to include them in a scoring technique. The very high rate of inconclusive evaluations produced by the field scoring technique indicates that in many cases the differences in responsivity between relevant and control questions were very small. This result is problematic from the point of view of the theory behind the CQT (i.e., that the control questions capture the psychological set of the truthful subject and elicit stronger reactions than the relevant questions).

The extreme base rate in our sample raises the possibility that the successes of the polygraphers were due only to the base rate, for by classifying all subjects as innocents the polygraphers could have guaranteed an accuracy rate of 13 out of 15. It must be remembered, however, that the polygraphers were not informed about the base rates. Furthermore, following Lykken's (1979) advice we reported the results separately for guilty and for innocent subjects. The accuracy rate was similar for both categories of subjects, and in fact was slightly higher for the smaller (i.e., guilty) category.

Finally, the proposed methodology raises an ethical issue. Is it fair to use subjects as was done here? Unfortunately, we do not believe that polygraphy can be validated in a controlled experiment without resorting to this kind of deception, because of the inherent incompatibility of the laboratory situation and real life. Thousands of people are interrogated yearly by the polygraph (in most cases by the CQT), and important decisions are based on the results of such testing. Yet the validity of this tool is not known, and the method of interrogation is highly controversial.

This study is a far cry from being the last word on assessing polygraph validity, but with large sample sizes the design it proposes can be used to improve these assessments.

---

[1] This alternative explanation was suggested by an anonymous reviewer.

Reference Note

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

References

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

Bersh, P. J. A validation of polygraph examiner judgments. *Journal of Applied Psychology*, 1969, *53*, 388–403.

Dawson, M. E. Physiological detection of deception: Measurement of responses to questions and answers during countermeasure maneuvers. *Psychophysiology*, 1980, *17*, 8–17.

Horvarth, F. S. The effects of selected variables on interpretation of polygraph records. *Journal of Applied Psychology*, 1977, *62*, 127–136.

Horvarth, F. S., & Reid, J. E. The reliability of polygraph examiner diagnoses of truth and deception. *Journal of Criminal Law, Criminology and Police Science*, 1971, *62*, 276–281.

Lykken, D. T. Psychology and the lie detection industry. *American Psychologist*, 1974, *29*, 725–739.

Lykken, D. T. The psychopath and the lie detector. *Psychophysiology*, 1978, *15*, 137–142.

Lykken, D. T. The detection of deception. *Psychological Bulletin*, 1979, *86*, 47–53.

Podlesny, J. A., & Raskin, D. C. Physiological measures and the detection of deception. *Psychological Bulletin*, 1977, *84*, 782–799.

Podlesny, J. A., & Raskin, D. C. Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology*, 1978, *15*, 344–359.

Raskin, D. C., & Hare, R. D. Psychopathy and detection of deception in a prison population. *Psychophysiology*, 1978, 126–136.

Raskin, D. C., & Podlesny, J. A. Truth and deception: A reply to Lykken. *Psychological Bulletin*, 1979, *86*, 54–59.

Reid, J. E., & Inbau, F. E. *Truth and deception: The polygraph ("lie-detection") technique.* Baltimore, Md.: Williams & Wilkins, 1966.

Received April 20, 1981 ■