Law and Human Behavior, Vol. 12, No. 1, 1988

Meta-analysis of Mock Crime Studies of the Control Question Polygraph Technique

John C. Kircher, Steven W. Horowitz, and David C. Raskin*

A review of results obtained from standard guilty and innocent treatment conditions in 14 mock crime studies of the control question polygraph technique revealed accuracies ranging from chance to 100% correct. The present study examined several factors that may have contributed to the observed variability in detection rates across studies. Those included sampling error, differences in the populations from which subjects were drawn (Subjects), differences in the nature of incentives provided to subjects for passing the polygraph test (Incentives), and differences in the methods for diagnosing truth or deception (Decision Policy). A meta-analysis revealed that approximately 24% of the variance in detection rates could be attributed to sampling error, and detection rates were correlated with types of Subjects ($r = .61$), Incentives ($r = .73$), and Decision Policies ($r = .67$). The highest diagnostic accuracies were obtained from nonstudent subject samples, when both guilty and innocent subjects were offered monetary incentives to convince the examiner of their innocence, and when conventional field methods were used for interpreting the physiological recordings and diagnosing truth and deception. Together, differences in Subjects, Incentives, and Decision Policies may account for as much as 65% of the observed variance in detection rates. The present findings highlight the importance of conducting mock crime experiments that closely approximate field conditions.

INTRODUCTION

In reviewing the literature on the accuracy of field polygraph techniques, the Office of Technology Assessment (OTA) (1983) summarized the results of 14 mock crime experiments. They reported that accuracy rates obtained under laboratory conditions were generally greater than chance, but there was considerable variability in the accuracy rates obtained by different investigators. Accuracy of decisions on subjects who were guilty of mock crimes ranged from a low of 71% correct (Szucko & Kleinmuntz, 1981) to a high of 100% correct (Dawson, 1980; Ginton, Netzer, Elaad, & Ben-Shakhar, 1982; Raskin & Hare, 1978). Even

* University of Utah.

0147-7307/88/0300-0079$06.00/0

greater variation occurred for innocent subjects. Accuracy of decisions on innocent subjects ranged from 49% correct (Szucko & Kleinmuntz, 1981) to 97% correct (Kircher & Raskin, 1982). Furthermore, the percentage of subjects correctly identified as truthful or deceptive in the laboratory experiments (61%) was more than 20% lower than that obtained from actual criminal suspects in the field studies (82%) (OTA, 1983). The present study explored the possibility that the observed differences in detection rates may be related to differences in the extent to which the research paradigms employed by different investigators were representative of field conditions.

In the typical mock crime experiment, subjects are randomly assigned to guilty and innocent treatment conditions. Subjects in the guilty condition commit a mock crime, such as the theft of an object of value from a place that the subject ordinarily would not frequent. Innocent subjects are given a general description of the crime but do not enact it. Subjects in both conditions are told to deny having committed the theft. They may be promised a reward if they can convince the polygraph examiner of their innocence, or they may be threatened with punishment if they cannot. After acting out the instructions, the subject is given a polygraph examination by an experimenter who is blind with respect to the subject's guilt or innocence.

The mock crime paradigm overcomes many of the problems and limitations of field research on polygraph techniques (Podlesny & Raskin, 1977). As compared to field settings, laboratory environments offer greater control over extraneous variables, testing contexts, instrumentation, and the qualifications and expertise of polygraph examiners. Since the subjects in a laboratory experiment are assigned to guilty and innocent treatment conditions, the accuracy of the polygraph technique may be assessed by comparing the test outcomes to the actual truthful or deceptive status of subjects (ground truth).

In field studies ground truth is rarely known with certainty. Inferences drawn from confessions or physical evidence obtained subsequent to field polygraph examinations may be used as criteria against which polygraph outcomes may be compared, but the validity of such criteria is open to question (Raskin, in press). Since field studies employ fallible criteria for establishing the veracity of criminal suspects in lieu of ground truth, they generally have less criterion validity than do laboratory experiments (OTA, 1983).

Although laboratory paradigms offer many advantages over field research, the accuracies obtained in mock crime experiments may not be representative of the accuracies obtained in the field (Lykken, 1981; Podlesny & Raskin, 1977). There may be important differences between individuals who agree to participate in psychological experiments and those who submit to polygraph examinations during criminal investigations. The consequences of failing the polygraph examination and the motivations of subjects to appear truthful on the test are typically greater in the field than in the laboratory. There may be important differences between the laboratory and field in terms of the amount of experience and the qualifications of those who administer the tests, interpret the physiological recordings, and render diagnoses of truth and deception.

The present study examined three factors that may affect the generalizability of results obtained from mock crime experiments. One variable concerned differences in the populations from which the subjects were drawn. Most mock crime experiments have used college students as subjects. College students constitute a relatively homogeneous group with respect to age, intelligence, educational background, socioeconomic status, and level of socialization. Compared to the general public, college students are more familiar with the academic settings in which mock crime experiments usually are conducted. Students may have some general understanding of the goals of behavioral research and may feel more comfortable playing the role of an experimental subject. The artificial nature of the mock crime paradigm may be more evident to the student than in the nonstudent subject. Such perceptions may reduce the subject's personal involvement in the outcome of the test and produce a psychological context that is substantially different from that which surrounds the polygraph examination of a person who is suspected of committing a criminal act.

In an attempt to obtain samples that are more representative of the population criminal suspects, some experimenters have recruited subjects from the community (e.g., Podlesny & Raskin, 1978; Rovner, Raskin, & Kircher, 1978). Others have sampled from the target population of individuals who commit crimes, such as psychopathic and nonpsychopathic prison inmates (Raskin & Hare, 1978) or psychopathic exoffenders (Hammond, 1980).

The subject's motivation to appear truthful on the polygraph test may also play a role in the outcome. Guilty and innocent subjects who undergo polygraph examinations in actual criminal cases are highly motivated to convince the polygraph examiner of their innocence. Depending on the circumstances, a deceptive polygraph outcome may result in the loss of prestige, a job, money, or even arrest or imprisonment. It is generally agreed that the typical mock crime paradigm does not completely simulate the qualitative and quantitative aspects of the motivational structure of the typical field polygraph examination (Lykken, 1981; Podlesny & Raskin, 1977).

The closest approximation to a realistic situation was achieved in a study conducted in Israel. Ginton et al. (1982) administered required aptitude tests to 21 Israeli policemen. The policemen were permitted to score their own tests, which gave them an opportunity to alter their answers. Unknown to the policemen, the answer sheets had been chemically treated so that it was possible to determine who had actually cheated on the test. Subsequently, the policemen were told that they were suspected of cheating and were asked to take a polygraph test. They were also led to believe that their professional careers might depend on the outcome of the test. Of the 15 policeman who agreed to be tested, 2 had actually cheated on the test. Since the police officers were unaware that they were subjects in an experiment, it is reasonable to assume that the investigators succeeded in creating a realistic motivational context for the polygraph examinations.

The least realistic incentives for passing the polygraph test were used by Szucko and Kleinmuntz (1981). They simply told the psychology undergraduate volunteers who served as subjects that "intelligent and well-adjusted" individuals can pass the test without being found guilty. Bradley and Ainsworth (1984) offered guilty subjects a $1.00 cash incentive to pass the test, but no attempt was made to motivate the innocent subjects. Barland and Raskin (1975) threatened college students in the innocent condition with the loss of college credit if they failed the test and promised guilty subjects a $10 bonus for passing the test.

Bradley and Janisse (1981) threatened guilty and innocent subjects with a "painful but not permanently damaging electric shock" if they failed the test. However, comparisons between their threatened and nonthreatened control subjects revealed that the threat of punishment had no effect on detection accuracy. In the remaining studies, guilty and innocent subjects were offered some type of a reward such as college credit (e.g., Honts, Hodes, & Raskin, 1985) or a monetary bonus (e.g., Podlesny & Raskin, 1978) for producing a truthful outcome on the test.

The amount of physiological data provided to the polygraph interpreters for making diagnoses also varied across experiments. In an attempt to control for the amount of data provided to the interpreter, the OTA examined the accuracy of judgments based on a maximum of three charts of physiological data for each subject. However, in one experiment (Szucko & Kleinmuntz, 1981), judgments of truth and deception were based on only one chart, and in other studies (e.g., Kircher & Raskin, 1982a) examiner judgments were based on as many as five charts. Although there may be some justification for attempting to standardize the amount of data provided to the polygraph examiners, decisions in field settings are not always based on three or fewer charts, and there is no requirement that decisions be reached in every case. In a typical field polygraph test, the examiner presents the series of test questions three times, evaluates the first three charts of data, and attempts to make a decision. If a decision cannot be reached at that point, one or two additional charts may be obtained and evaluated. If the polygraph examiner is unable to reach a diagnosis after evaluating as many as five charts of data, the test is considered inconclusive.

Instead of controlling the number of polygraph charts evaluated by the polygraph interpreters, the present study explicitly considered the extent to which the number of charts provided to the interpreters fulfilled the requirements of standard field practice. Each experiment was categorized according to whether or not the methods of chart interpretation and decision rules employed by practicing field examiners were accurately represented in the experiment. That procedure used all of the available physiological data and permitted an examination of the effects on accuracy rates attributable to violations of conventional methods of chart interpretation and decision rules.

The classification strategy described above is confounded with another variable, the effects of which cannot be adequately assessed with the available data. Field polygraph examiners use one of two general methods for diagnosing truth and deception. In the oldest approach, the polygraph examiner forms a global impression of the subject's physiological responses to test questions (Reid & Inbau, 1966). To reach an overall determination of truth or deception, that information is combined in some unspecified manner with evaluations of the case facts and the subject's demeanor during the test.

The other general diagnostic approach is known as numerical scoring

(Raskin, 1982). The numerical method attempts to minimize the influence of extrapolygraphic sources of information on the decision maker and to maximize the reliability of examiner judgments. Physiological responses to test questions are systematically scored, the obtained scores are summed, and the subject is classified as truthful, deceptive, or inconclusive by comparing the total numerical score to standard criteria.

Global evaluations of the polygraph charts were performed in two mock crime experiments (Ginton et al., 1982; Szucko & Kleinmuntz, 1981). However, the Reid-trained examiners and student-trainees in the Szucko and Kleinmuntz study were not provided with the nonphysiological sources of information on which they had been trained to rely; their decisions were based on only one chart of data rather than three or more charts, and they were required to render a definite decision in every case. As a consequence, the use of global evaluators in the Szucko and Kleinmuntz study was confounded with the use of arbitrary decision rules.

Both global and numerical evaluations of the physiological data were performed in the study by Ginton et al. (1982). However, from their description it is not clear whether the polygraph examiners had been trained in global or numerical methods of evaluation, or both. Also, of the 15 subjects who participated in their experiment, only two were guilty of cheating. In view of the limited number of subjects in the Ginton et al. study and the constraints placed on the inadequately trained polygraph examiners in the Szucko and Kleinmuntz study, these two studies do not clearly represent outcomes obtained by global methods of evaluation. A direct comparison of the accuracies of global and numerical interpreters in a field study may be found in Raskin, Barland, and Podlesny (1978).

The OTA study found considerable variability in reported levels of diagnostic accuracy across studies, but it made no attempt to analyze that variability. The present study used procedures described by Hunter, Schmidt, and Jackson (1982) and Glass (1976) to perform a meta-analysis of the observed variability in detection rates. According to Hunter et al., much of the variance in results obtained by different studies may be attributed to statistical artifacts such as sampling error, differences in the reliability of measurement and the range of independent variables, and computational and typographical errors. Various applications of their techniques in the area of personnel selection revealed that the first three artifacts accounted for 72% of the variance in research findings, and approximately 60% of the total variance could be explained by sampling error alone (Schmidt & Hunter, 1981).

Hunter et al. suggested that if more than 75% of the variance in research findings is due to the effects of sampling error, errors of measurement, and range restriction, then the search for substantive differences among the studies (moderator variables) is unwarranted. However, the present meta-analysis assessed only the effects of sampling error because the available information was insufficient to assess errors of measurement, and the range of the independent variable was held constant by limiting the analysis to standard guilty and innocent treatment conditions, as discussed below. Since only sampling error was considered, a modification of the 75% decision rule seemed appropriate. Following suggestions by

Peters, Hartke, and Pohlmann (1986), we decided not to search for moderator variables if sampling error accounted for more than 60% of the observed variance in detection rates among the 14 studies.

METHOD

Literature Base and Case Selection

Sixteen mock crime studies of the control question technique were found in the literature. Two of those studies were omitted from the present analysis because an index of diagnostic validity could not be computed when only guilty (Widacki & Horvath, 1978) or only innocent subjects (Heckel, Brokaw, Salzberg, & Wiggins, 1962) participated in the experiment.

Some of the variance in the detection rates reported in the OTA study may be attributed to effects of experimental treatments (e.g., training in the use of physical countermeasures) that had been implemented in some experiments but not in others. In the present study, that source of variance was removed by limiting the analysis to control subjects and to subjects who had received experimental treatments that had no significant effect on the accuracy of diagnoses. That requirement resulted in the loss of 93 (11%) of the total number of 858 subjects who had participated in the 14 experiments. However, it substantively corrected the range of treatments to standard guilty and innocent control conditions.

Assessments of Diagnostic Accuracy

An index of diagnostic accuracy was obtained for each study by correlating the judgments by the polygraph interpreters (coded as 1 for truthful decisions, 0 for inconclusive, and −1 for deceptive decisions) with the criterion of guilt or innocence (coded as 1 for innocent subjects and −1 for guilty subjects). An obtained correlation of 0.0 would indicate that there was no relationship between the judgments made by the polygraph interpreter and the criterion, and a correlation of 1.0 wound indicate that the judgments of truth and deception were perfectly accurate.

Our use of the correlation coefficient is based on the assumption that there is an underlying order to the polygraph interpreters' judgments, with inconclusive outcomes being treated as intermediate values along a truthful/deceptive continuum. Although inconclusive outcomes may be viewed as failures of the technique and their occurrence would reduce the value of the correlation coefficient, they would not be weighted as heavily as false positive or false negative decision errors. Thus, the correlation coefficient provides a measure of detection efficiency that is consistent with the real-world consequences of various types of polygraph outcomes. Furthermore, procedures for performing a meta-analysis were originally developed for analyzing variability among correlation coefficients, and the present method for measuring detection efficiency facilitated their application.

RESULTS

The outcomes obtained from guilty and innocent subjects, the sample sizes, and the obtained correlation between the judgments by the polygraph interpreter and the criterion (r) are summarized for each of the 14 experiments in the last column of Table 1. A wide range of correlation coefficients was obtained.

The estimate population variance of the correlations is given by $Sr^2 = \Sigma n_i (r - \bar{r})^2/n_i$, where n_i is the sample size and r_i is the observed r for ith study. The average of squared deviations between the observed r's and $\bar{r}$ weighted by their respective sample sizes was .0245. The estimated variance due to sampling error was obtained by $Se^2 = k(1 - \bar{r}^2)^2/N$, where k is the number of studies and N is the total number of subjects (Hunter et al., 1982). The variance in observed correlations due to sampling error was .0058. Thus, only 23.8% (.0058/.0247) of the observed variability in detection efficiency was due to sampling error. Since that is considerably lower than the criterion value of 60%, a search for moderator variables was appropriate.

Three dichotomous measures were developed to reflect the extent to which investigators employed methods that were representative of existing field conditions. Similar to the meta-analytic techniques used by Smith and Glass (1977), for each study a score of 0 or 1 was assigned for each characteristic to indicate a relatively low or high degree of generalizability to the field situation. As illustrated in Table 2, a score of 0 on the Subject dimension indicated that the subjects were college students or student actors, and a score of 1 indicated that subjects were not students. A score of 0 on Incentives indicated that minimal incentives for producing a truthful outcome on the test were provided to guilty and/or innocent subjects, and a score of 1 indicated that stronger and equal incentives were provided to both groups. A score of 0 on Decision Policy was assigned when diagnoses of truth and deception were based on nonstandard field scoring techniques, and a score of 1 was assigned when standard field methods were used, as previously described.

The extent to which limitations on the generalizability of laboratory results may be related to detection rates was assessed by correlating the scores on each of the three design characteristics with the obtained correlations between the interpreters' diagnoses and the criterion (r's). Correlations were also obtained between the scores on the three design characteristics and z-score transformations of the obtained r's. The results obtained with z scores were uniformly stronger than those obtained using the correlations, but the same pattern of results emerged. For ease of interpretation, only the results obtained with correlation coefficients are reported.

To account for differences in sample size, weighted correlations were obtained according to procedures outlined by Hunter et al. (1982). They are shown in Table 3.

The correlations between each of the three design characteristics and detection efficiency are shown in the first column. In each case, a substantial positive and significant relationship was observed. The more closely the subject sample resembled the field population, the more accurate were the decisions. Relatively

Table 1. Percent Outcomes Obtained under Standard Guilty and Innocent Conditions

	Guilty (n = 382)				Innocent (n = 383)				
Study	n	Correct	Wrong	Inconclusive	n	Correct	Wrong	Inconclusive	Detection Efficiency
Barland and Raskin (1975)	36	64	8	28	36	42	17	42	.51
Bradley and Ainsworth (1984)[a]	16	88	13	0	8	75	13	13	.69
Bradley and Janisse (1981)	96	60	14	26	96	58	9	32	.57
Dawson (1980)[b]	12	100	0	0	12	75	8	17	.83
Gatchel et al. (1984)	14	50	7	43	14	79	0	21	.76
Ginton et al. (1982)	2	100	0	0	13	85	15	0	.65
Hammond (1980)	32	72	3	25	30	40	20	40	.57
Honts et al. (1983)[c]	10	80	0	20	10	70	20	10	.71
Honts et al. (1985)[d]	31	77	3	19	31	45	19	35	.61
Kircher and Raskin (1982a)	50	88	6	6	50	86	6	8	.84
Podlesny and Raskin (1978)	20	70	15	15	20	90	5	5	.75
Raskin and Hare (1978)	24	88	0	12	24	88	8	4	.87
Rovner et al. (1979)[e]	24	88	0	12	24	88	8	4	.87
Szucko and Kleinmuntz (1981)	15	71	29	0	15	49	51	0	.21
Weighted means		74	8	18		66	12	22	.66

[a] Sixteen intoxicated guilty subjects excluded.
[b] Data from delayed answer excluded.
[c] Ten countermeasure-trained subjects excluded.
[d] Forty-three countermeasure-trained subjects excluded.
[e] Twenty-four countermeasure-trained subjects excluded.

Table 2. Characteristics of Mock Crime Experiments

Study	Characteristic (assigned code)		
	Subject sample	Incentives	Decision policy
Barland and Raskin (1975)	(0) College students	(0) Course credit and $10 bonus for guilty only	(0) 3 charts
Bradley and Ainsworth (1984)	(0) College students	(0) $1 bonus for guilty only	(0) 3 charts and modified scoring
Bradley and Janisse (1981)	(0) College students	(0) Course credit or threat of electric shock	(0) 3 Charts
Dawson (1980)	(0) Student actors	(1) $5 pay + $5 bonus	(0) 2 charts
Gatchel et al. (1984)	(0) Medical students and staff	(1) $15 pay + $10 bonus	(0) 3 charts
Ginton et al. (1982)	(1) Policemen	(1) Career threat	(1) Field technique
Hammond (1980)	(1) Students, alcoholics and exoffenders	(1) $7 pay + $10 bonus	(0) 2 charts
Honts et al. (1983)	(0) College students	(1) Course credit + $15 bonus	(1) Field technique
Honts et al. (1985)	(0) College students	(0) Course credit	(1) Field technique
Kircher and Raskin (1982a)	(1) General community	(1) $8 pay + $17 bonus	(1) Field technique
Podlesny and Raskin (1978)	(1) General community	(1) $5 pay + $10 bonus	(1) Field technique
Raskin & Hare (1978)	(1) Psychopathic and nonpsychopathic prisoners	(1) $20 bonus	(1) Field technique
Rovner et al. (1978)	(1) General community	(1) $7.50 pay + $10 bonus	(1) Field technique
Szucko and Kleinmuntz (1981)	(0) College students	(0) Threat to self esteem	(0) 1 chart and no inconclusives

low levels of detection efficiency were obtained in studies with college student subjects, and the highest levels of detection efficiency were obtained from more heterogeneous samples of subjects, including psychopathic and nonpsychopathic prison inmates and exoffenders.

Strong correlations with detection efficiency were also obtained for the Incentives and Decision Policy variables. Studies in which both guilty and innocent subjects were offered monetary incentives for a truthful outcome on the polygraph test produced higher decision accuracies than those that did not. In addi-

Table 3. Correlations among Study Characteristics and Detection Efficiency

	Detection efficiency	Subject sample	Incentives
Subject sample	.61		
Incentives	.73	.83	
Decision policy	.67	.62	.55

tion, accuracy of decisions was greatest when trained and experienced polygraph examiners evaluated three or more charts of recorded physiological data using standard numerical scoring criteria and decision rules. Nonstandard scoring techniques and the arbitrary decision criteria employed by Szucko and Kleinmuntz (1981) were associated with the lowest levels of detection efficiency.

A multiple regression analysis was performed to assess the proportion of variance in detection efficiency that may be attributed to the combined effects of the three design characteristics. The three dichotomously coded design variables were simultaneously entered into the regression equation to predict the observed correlation between interpreter judgments and the criterion of guilt and innocence. The analysis produced an R^2 of .65, which suggests that the Subject, Incentive, and Decision Policy variables may account for as much as 65% of the observed variance in detection rates.

DISCUSSION

The present findings suggest that diagnostic accuracy in mock crime experiments depends on the extent to which the subjects, incentives, and procedures for evaluating the physiological data are representative of field conditions. Those factors may account for much of the variance in the accuracies obtained in laboratory experiments, and they may account for the discrepancy between accuracy rates in laboratory and field studies of the control question technique. However, one cannot infer that differences along any of the three dimensions examined in the present study were causally related to detection efficiency. Although the present findings are suggestive of such relationships, they are not definitive since no attempt was made to manipulate the number and/or types of threats to the generalizability of laboratory results.

The adequacy of the criteria that we used to rate the procedures employed by different investigators may be questioned. This issue is particularly important in light of the small number of studies on which the present findings were based. Under these circumstances, even small changes in the criteria could have large effects on the correlations with detection rates.

The observed correlation between detection rates and decision policies highlights the importance of using standard field scoring techniques in laboratory experiments whenever inferences are to be drawn about the accuracy of such techniques in field settings. The correlations of detection rates with subject characteristics and incentives are of greater theoretical interest. Significantly lower detection rates were obtained from college student as compared to nonstudent samples and when minimal negative consequences were associated with a deceptive polygraph outcome. Relative to other members of the community, college students may be more familiar with the nature and objectives of psychological experiments, feel more comfortable in research settings, experience less emotional arousal while performing their tasks, and have little invested in the outcome of the test. Furthermore, personal involvement in the task may be especially difficult to achieve with college students. For a college student, the loss of a $20 bonus for failing the polygraph test may be unimportant; but to the unemployed or prison inmates with limited resources and opportunities to earn money, $20 may be a significant loss. Experiments are needed to assess the effects of personal involvement on detectability. The results of such studies may explain some of the variance in the results of laboratory studies and facilitate attempts to develop a comprehensive theory of detection of deception.

Given the limitations of mock crime analogs in which subjects are informed as to their roles as experimental subjects, a method for assessing the adequacy of the mock crime paradigm is badly needed. Highly realistic procedures that involve entrapment and deception of subjects (Ginton et al., 1982) are not likely to gain widespread acceptance among researchers in this area, since they might violate the ethical standards of research on human subjects (American Psychological Association, 1981). A possible solution would compare sets of intercorrelations among components of physiological responses observed under laboratory and field conditions. Computer techniques for data quantification and multivariate statistical methods, such as confirmatory factor analysis, multivariate analysis of variance, and discriminant analysis, may be used for those purposes. Use of such techniques might reveal whether or not the contextual and motivational components of field polygraph examinations that are difficult to simulate in the laboratory are important determinants of subjects' physiological responses to test questions. Similar techniques have already been used to provide powerful tests of quantitative and qualitative differences between the patterns of physiological activation that accompany truthfulness and deception during control question polygraph tests (Kircher & Raskin, 1982b), and their application in this area might prove to be fruitful.

REFERENCES

American Psychological Association (1981). *Ethical Principles of Psychologists*. Washington, D.C.

Ben-Shakhar, G., Lieblich, I., & Kugelmass, S. (1975). Detection of information and GSR habituation: An attempt to derive detection efficiency from two habituation curves. *Psychophysiology, 12*, 283–288.

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

Bradley, M. T., & Ainsworth, D. (1984). Alcohol and psychophysiological detection of deception. *Psychophysiology, 21*, 63–71.

Bradley, M. T., & Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and pupillary measures, *Psychophysiology, 18*, 307–315.

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

Ekman, P. (1985). *Telling lies: Clues to deceit in the market place, politics and marriage*. New York: W. W. Norton & Co.

Gatchel, R. J., Smith, J. E., & Kaplan, N. M. (1984). The effect of propranolol on polygraphic detection of deception. Unpublished manuscript.

Ginton, A., Netzer, D., & Elaad, E., & Ben-Shakhar, G. (1982). A method for evaluating the use of the polygraph in a real-life situation. *Journal of Applied Psychology, 67*, 131–137.

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Research, 5,* 3–8.

Hammond, D. L. (1980). The responding of normals, alcoholics and psychopaths in a laboratory lie-detection experiment. Unpublished doctoral dissertation, California School of Professional Psychology.

Heckel, R. V., Brokaw, J. R., Salzberg, H. C., & Wiggins, S. L. (1962). Polygraphic variations in reactivity between delusional, nondelusional, and control groups in a crime situation. *Journal of Criminal Law, Criminology and Police Science, 53,* 380–383.

Honts, C. R., Hodes, R. L., & Raskin, D. C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology, 70,* 177–187.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1983). Detection of deception: Effectiveness of physical countermeasures under high motivation conditions. *Psychophysiology, 20,* 446–447.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, California: Sage.

Kircher, J. C., & Raskin, D. C. (1982a). Cross-validation of a computerized diagnostic procedure for detection of deception. *Psychophysiology, 19,* 568–569 (Abstract).

Kircher, J. C., & Raskin, D. C. (1982b). Is there a "specific lie pattern" of autonomic responses? *Psychophysiology, 19,* 569 (Abstract).

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

Lykken, D. T. (1981). *A Tremor in the Blood,* New York: McGraw-Hill.

Office of Technology Assessment (1983). *Scientific validity of polygraph testing: A research review and evaluation.* OTA-TM-H-15. Washington, DC: U.S. Government Printing Office.

Peters, L. H., Hartke, D. D., & Polhman, J. T. (1986). Fiedler's contingency theory of leadership: An application of the meta-analysis procedures of Schmidt and Hunter. *Psychological Bulletin, 97,* 274–285.

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

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

Raskin, D. C. (1979). Orienting and defensive reflexes in the detection of deception. In H. D. Kimmel, E. H. van Olst, & J. F. Orlebeke (Eds.), *The orienting reflex in humans,* Hillsdale, New Jersey: Lawrence Erlbaum, pp. 587–605.

Raskin, D. C. (1982). The scientific basis of polygraph techniques and their uses in the judicial process. In: A. Trankell (Ed.), *Reconstructing the past.* Deventer, the Netherlands: Kluwer.

Raskin, D. C. (in press). Methodological issues in estimating polygraph accuracy in field applications. *Canadian Journal of Behavioural Science.*

Raskin, D. C., Barland, G. H., & Podlesney, J. A. (1978). *Validity and Reliability of Detection of Deception.* (Stock Number 027-000-09692-2) Washington, DC: U.S. Government Printing Office.

Raskin, D. C., & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology, 15,* 126–136.

Reid, J. E., & Inbau, F. E. (1966). *Truth and deception: The polygraph technique.* Baltimore: The Williams & Wilkins Co.

Rovner, L. I., Raskin, D. C., & Kircher, J. C. (1978). Effects of information and practice on detection of deception. *Psychophysiology, 16,* 198 (Abstract).

Schmidt, F., & Hunter, J. (1981). Employment testing: Old theories and new research findings. *American Psychologist, 36,* 1128–1137.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32,* 752–760.

Szucko, J. J., & Kleinmuntz, D. (1981). Statistical versus clinical lie detection. *American Psychologist, 36,* 488–496.

Widacki, J., & Horvath, F. (1978). An experimental investigation of the relative validity and utility of the polygraph technique and three other common methods of criminal investigation. *Journal of Forensic Sciences, 23,* 596–601.