

THE ACCURACY AND UTILITY
OF POLYGRAPH TESTING

DEPARTMENT OF DEFENSE
WASHINGTON, D.C.
1983

CONTENTS

Foreword

Overview

Chapter 1 - Research on Polygraph Criterion Validity and Reliability

Field Studies of Criterion Validity

Relationship of Blind Examiners and Ground Truth

Relationship of Examiner Decision, Ground Truth Unknown

Laboratory Studies

Control-Question Test, Criterion Validity

Control-Question, Reliability

Relevant-Irrelevant Technique, Criterion Validity

Relevant-Irrelevant, Reliability

Guilty Information Tests, Criterion Validity

Guilty Information Tests, Reliability

Chapter 2 - Analysis of Research

Analysis

The Bersh Field Validation Study

Experience of Professional Investigators
and Quality Control Personnel

Laboratory Experiments with Mock Crimes

Summary

Chapter 3 - Utilization of the Polygraph

The Polygraph in the Federal Government

Federal Agencies

Polygraph Examiners in the Department of Defense

Number of Examiners (Table)

Annual Polygraph Utilization (Table)

Criminal Investigation

History

Polygraph as an Aid to Investigation

Tables on Use in Criminal Investigations

Criminal Investigation Case Examples

Criminal Conduct Revealed During Screening Examinations

Intelligence, Counterintelligence and Security

History

Specific Investigations of Espionage

Examples of Polygraph Cases - Espionage

Suitability - Statistics

Polygraph and the Traditional Methods

1953 Study of Polygraph/Background
1983 Study of Polygraph/Background
1981 Polygraph Results from Cleared Military Personnel
1974 - 1979 Polygraph Results from Cleared Military Personnel
1983 Study of Interviews/Polygraph
Use of the Polygraph in Army Personnel Security Cases

Surveys of Examinees

1982 Air Force Survey
1951 Survey of Atomic Energy Commission Employees
1956 Survey of NSA Employees
Surveys in Law Enforcement and Commerce

Quality Control

U.S. Army Polygraph Training Facility

Basic Polygraph Training
Advanced Training

Primary Polygraph Testing Techniques

Chapter 4 - Summaries of Research

Field Abstracts

Gordan H. Barland and David C. Raskin (1976)
Akiva Ben-Ishai (1962)
Philip J. Bersh (1969)
M. E. Bitterman and F. L. Marcuse (1947)
Avital Ginton et al. (1982)
Eugene C. Edel and Jacob Jacoby (1975)
Eitan Elaad and Esther Schahar (1976)
Robert H. Edwards (1981)
Frank Horvath (1977)
Frank Horvath and John E. Reid (1971)
Fred L. Hunter and Philip Ash (1973)
Verne W. Lyon (1936)
Jesse Orlansky (1962)
Robert B. Peters (1982)
David C. Raskin (1976)
Stanley M. Slowik and Joseph P. Buckley (1975)
Douglas E. Wicklander and Fred L. Hunter (1975)
Jan Widacki (1982)

Laboratory Abstracts: Control Question Techniques

Gordan H. Barland (1981)
Gordon H. Barland and David C. Raskin (1975)
M. T. Bradley and Michel P. Janisse (1981)
Robert J. Gatchel, John E. Smith, and Norman M. Kaplan (1983)
David L. Hammond (1980)

Charles R. Honts (1982)
D. C. Raskin and R. D. Hare (1978)
L. I. Rovner, D. C. Raskin and J. C. Kircher (1978)
William M. Waid, Emily C. Orne and Martin T. Orne (1981)
William M. Waid, Martin T. Orne and Stuart K. Wilson (1979)
Jan Widacki and Frank Horvath (1978)

Laboratory Abstracts: Relevant-Irrelevant Techniques

Richard H. Blum and William Osterloh (1968)
Eileen J. Correa and Henry E. Adams (1981)

Laboratory Abstracts: Guilty Knowledge and Peak of Tension Techniques

Kristan H. Balloun and David S. Holmes (1979)
P. O. Davidson (1968)
Miroslav Dufek (1969)
Jan Kronbergerova and Miroslav Dufek (1969)
David T. Lykken (1959)
David T. Lykken (1960)
Kazuo Ohnishi, et al. (1976)
Robert M. Stern, et al. (1981)
William M. Waid, et al. (1978)
William M. Waid, et al. (1981)
Kazunoba Yamaoka and Akihiro Suzuki (1980)

References Cited

Most of the references cited in the text will be found in the abstracts section. Those references which are not in the abstracts are listed after the abstracts.

Foreword

FOREWORD

This analysis of the scientific literature on the accuracy of the polygraph, with supporting information on use and utility, was prepared at the direction of Deputy Under Secretary of Defense Richard G. Stilwell, General, USA (Ret.). General Stilwell, with the concurrence of the Director of the National Security Agency, appointed Norman Ansley to lead a working group of senior examiners from the military services to prepare a report, setting forth the basis upon which the polygraph is utilized in the Department of Defense. In addition, the report was to include information on the use and results of polygraph testing in the Department.

The working group, listed below, provided information from their resources on the use of the polygraph and case examples. They also served as an editorial board, reviewing the text. This report was prepared by Norman Ansley, Chief, Polygraph Division, Office of Security, National Security Agency, and Marcia Garwood, Ph.D., Research and Special Branch, Polygraph Division, Office of Security, National Security Agency. Gordon H. Barland, Ph.D., served as a scientific editor of the report. William Fedor, Deputy Director (Personnel Security) Counterintelligence and Investigative Programs, and William H. Bell, Security Specialist, Counterintelligence and Investigative Programs Directorate, ODUSD (P), represented DoD in the development of this report.

THE WORKING GROUP

Norman Ansley
Chief, Polygraph Division
Office of Security
National Security Agency

Robert A. Brisentine, Jr.
Chief Polygraph Examiner
U.S. Army Law Enforcement
U.S. Army Criminal Investigation Command

Louise S. Fuse
Deputy Polygraph Program Manager
U.S. Army Intelligence and Security Command

B. Frank Bloomingburg
Special Assistant
Scientific Investigations
Headquarters, Naval Investigative Service

James E. Hardy
Chief, U.S. Air Force Polygraph Program
Headquarters, USAF Office of Special Investigation

This book was produced in less than 30 days. It is based on the material currently available. The tables on DoD use are based on the semiannual reports of the military services and NSA. An enlarged and final edition will be available later this year. It will contain more summaries, particularly foreign material. There will be additional topics addressed, and a more detailed analysis of the data.

OVERVIEW

OVERVIEW

The federal government has over sixty-five years of experience with polygraph techniques in criminal and counterintelligence testing. Beginning in 1951, the Department of Defense established a polygraph school for the training of military examiners. Today, that course is considered second to none. It conducts basic and advanced courses to train almost all of the polygraph examiners for the federal service.

The use of the polygraph for the investigation of crimes is a well established practice. Known errors in the field use are exceedingly rare. Where examinees are found to be deceptive during testing, the confession rate is consistently high, despite the fact that all of these criminal suspects had already been interrogated by an experienced investigator. Moreover, there is nothing about the polygraph technique that likely to cause a false confession because of the requisite low key questioning. The use of the polygraph to confirm admissions and confessions with additional testing also reduces the possibility of false confessions. This study includes a number of representative cases of criminal investigations in which the polygraph has been a significant factor.

The polygraph is extremely useful in intelligence and counterintelligence operations. This is evidence of the deterrent effect of screening examinations. Examples of espionage and attempted espionage cases detected by polygraph examinations are included in this study. Without the polygraph as an investigative tool, a number of espionage cases never would have been solved. Helmich, Kampiles, and Barnett probably would not have been successfully prosecuted without the skillful application of polygraph techniques. In addition, there is definite evidence that some extremely sensitive U.S. intelligence operations would have been penetrated by hostile intelligence services if the polygraph had not been employed in screening for clearance and access. Examiners conducting screening cases have

obtained confessions from applicants of their intention to commit espionage. In other cases they developed such significant admissions that penetration attempts by hostile intelligence were detected and neutralized. Screening has also kept our intelligence agencies from hiring some extremely undesirable people. Examiners, in FY 82, obtained admissions from applicants of undetected crimes involving murder, attempted murder, arson, rape, and numerous other felonies.

The polygraph field is one of those rare situations where the practice has outpaced the research. While scientists are debating how the various testing formats work, and are trying to devise a satisfactory theoretical framework, others are conducting research on the development of improved techniques and advanced instrumentation. Much of the literature which addresses the issue of polygraph validity was conducted for other purposes. Some of the research is very limited in scope, some is dated, and some is flawed in design. Despite these limitations they produce results significantly above chance. Those studies which approximate field conditions by using field instruments, trained and experienced examiners, standardized techniques, and a complete field pretest interview, produce higher rates of accuracy than those conducted with one physiological channel, make-shift or novel techniques, and untrained personnel.

The polygraph is not limited to our culture. It is significant that research results are similar throughout the world, regardless of varied cultures and languages.

It is important to realize that polygraph examinations are not conducted in isolation. Their use is always in the context of a larger program. They play a role in investigations but they are never a substitute for investigation. Polygraph examiners do not make recommendations or decisions in regard to further investigation or the adjudication of the results. In screening, the polygraph examination is not used alone for clearance and access. It is an adjunct to the

national agency check and background investigation. Decisions and adverse actions are not made solely on the results of an analysis of polygraph charts. All other factors in a situation are considered before decisions are made.

Used with prudence, and a full knowledge of its limitations, the polygraph will continue to play a role in our criminal justice system and counterintelligence operations.

Chapter 1

Research on Polygraph Criterion Validity and Reliability

RESEARCH ON POLYGRAPH CRITERION VALIDITY AND RELIABILITY

FIELD STUDIES

Field studies have investigated both the validity and reliability of the polygraph using various criteria. These studies typically use field model polygraphs and record respiration, electrodermal activity, and cardiovascular activity. Professional polygraph examiners are typically used, but they differ widely in experience. Some examiners are still in internship training programs; others have as much as 20 years of field experience. Field studies have used both relevant-irrelevant and control-question techniques. A description of the standardized formats appears in the Chapter on Utility.

The original examiner's decision of truth or deception is often a clinical judgment based on information from physiological recordings, subject behavior, and case facts, although some data on the numerical scoring of polygraph charts also will be discussed. Most of the field studies suffer from significant methodological problems.

Two types of studies will be discussed. First, studies investigating the relationship between the examiner's decision and some criterion of truth or deception will be presented as studies of polygraph criterion validity. All of these studies use the clinical diagnosis of the original examiner to predict the criteria. However, two of the studies (Barland and Raskin, 1978; Ginton, Daie, Elaad and Ben-Shakhar, 1982) also use both numerical scoring by the original examiners, and decisions of examiners not present during the test based on numerical evaluation and chart scrutiny to predict truth or deception.

Second, studies investigating the relationship between the decisions of a blind examiner and actual truth or deception using verified polygraph cases will be discussed. Some of these studies have been used as evidence for polygraph

criterion validity and there are problems with this interpretation. The final section focuses on the reliability of the polygraph examiner's decision.

FIELD STUDIES OF CRITERION VALIDITY

Field studies of criterion validity have used various criteria of ground truth. Determination of ground truth is the major problem for such studies. The ideal criterion is independent of the polygraph examiner decision of truth or deception and accurately reflects subject guilt or innocence. If examiner decision and the criteria are not independent, the accuracy of the polygraph in correctly classifying innocent and guilty subjects will be reported as spuriously high. However, most of the independent determinations of ground truth probably produce spuriously low correct classification rates due to their nature. There is only one study where the criterion is both independent and accurate (Ginton, et al. 1982). This study will be discussed in detail at the end of this section.

Several studies have used the criterion of case disposition to assess the accuracy of the examiners' decision (Lyon, 1936; Peters, 1982; Edwards, 1981). Case disposition as a criteria may be neither independent nor accurate. Lyon (1936) determined the dispositions of randomly-selected cases involving 100 juvenile criminal suspects (81 boys and 19 girls over age 15). Twenty polygraph tests were classified truthful and 80 were classified as deceptive. These cases were followed to determine if examiners' results were corroborated by positive evidence or confession. Only 40 cases could be verified. Seven cases classified as truthful by the examiner were verified as truthful by additional evidence. Thirty-three of the deceptive cases could be verified by confessions; some of the confessions occurred immediately after the test. Thus, the examiner had correctly classified all cases. Lyon, however, gives insufficient detail about the verification procedure. Testing involved a unique procedure but was in the relevant-irrelevant category. Respiration and cardiovascular activity were recorded.

Peters (1982) reported on the disposition of 220 cases involving stipulated polygraph examinations conducted at a state crime laboratory from 1976 to 1979. Stipulation means that the prosecution, defense, and judge agree to admit results of the polygraph examination into evidence with the examiner testifying and subject to cross-examination, although the examiner actually testified about the subject's truth or deception in only 11 cases. Control-question procedures and numerical chart analysis (with an inconclusive range) were used. Of the initial 220 examinations, 25 were eliminated because they were inconclusive, 23 cases were eliminated because the test could not be completed. The disposition of 100 cases indicated that the subject was innocent or telling the truth (95 criminal suspects had charges dropped, 1 criminal suspect was acquitted, and 4 victims or witnesses were involved in trials where the accused was convicted). The polygraph examiner had classified 88% of these cases as truthful. The disposition of 63 cases indicated the subject was guilty or lying (10 criminal suspects were convicted at trial, 51 pleaded guilty, and 2 alleged victims were involved in trials where the accused was acquitted). The polygraph examiner had classified 98% of these cases as deceptive. Six cases were pending and 3 cases had charges dropped with conditions (these cases are not included in the correct classification figures).

Edwards (1981) surveyed licensed Virginia polygraph examiners to determine the correspondence between polygraph examiner decisions and dispositions of case in specific issue examinations. Only results from law enforcement examiners will be discussed; the response rate for private firms was so low that it is meaningless. Of the 2,433 polygraph examinations, 959 (594 guilty and 365 innocent) were verified by means such as acquittal, conviction, confession, etc. The examiners had correctly classified 99% of the guilty cases and 97.5% of the innocent cases.

Lack of independence between examiner decision and the criterion can occur in several ways and it is impossible to determine the extent the nonindependence influences the percentage of correct classifications. The examiner can influence

the trial process either by increasing the probability of conviction or that charges are dropped. This is a particular problem in the Peters study involving stipulated polygraph tests. Peters does mention that the examiner testified about subject guilt or innocence in only 11 cases, but it is unclear the extent to which the results of the polygraph test were influential in the decision to go to trial or drop charges. Confession also is not independent of the polygraph examination. Examiners usually interrogate only those showing physiological responses to relevant questions. Thus, those showing no deceptive responses will be under less pressure to confess. It also is possible to induce a false confession, although that is extremely rare. Some researchers think the use of confession invalidates the study, other think confession is acceptable as a criterion (see Horvath, 1971; Lykken, 1981; Raskin and Podlesny, 1979).

Field studies using case disposition also are problematical because they report data on a restricted sample. Typically, a large number of cases cannot be verified and it is not clear if verified cases differ from unverified cases in the ease with which examiner can make a decision about the case. Horvath (1977) provides preliminary data that verified and non-verified cases may be quite similar.

Two field studies have used panel judgments as the criterion by which to validate polygraph examiner judgment (Bersh, 1969; Barland and Raskin, 1976). Barland and Raskin (1976) is one of the studies Lykken (1979; 1981) cites in maintaining that polygraph tests have a high rate of false positive errors.

Bersh (1969) investigated the relationship between examiner judgment in criminal investigations by the military services and unanimous guilt-innocence decisions by a panel of four Judge Advocate General Attorneys who had complete access to case files except for the polygraph examinations. Examiners also had access to a subject's case file, but the author maintained the file was not complete and that polygraph tests are only given when there is doubt about a person's guilt or innocence. Cases were selected at random from 1963 to 1966 with some

restrictions. Cases involving confession by the subject and inconclusive judgments by the examiner or the panel were eliminated. Cases were selected so that there were similar numbers of deception and no deception examiner judgments and a mix of zone comparison tests, Modified General Question Tests and relevant-irrelevant tests. The first two formats are control question tests. However, Berish grouped the Modified General Question Test and the relevant-irrelevant test in reporting the data. This was referred to as the General Question Test. Decisions by the panel were unanimous in 157 cases. With the General Question Test group, the examiner classified as deceptive 97% of the cases the panel judged as guilty, and classified as truthful 90% of the cases the panel judged as innocent. With the zone-comparison test, the examiner agreed with 90% of the cases the panel judged as guilty and 94% of the cases the panel judged as innocent.

Barland and Raskin (1976) investigated the correspondence between polygraph examiner decisions and panel decision (1 judge, 2 prosecuting attorneys, 2 defense attorneys) for 92 sequential independent cases involving criminal suspects. Panel decisions were based on evidence collected by untrained investigators and the panel had no knowledge of the outcome of the polygraph examinations. Polygraph examinations used control-question procedures. Panel decision was based on a numerical score by each panel member. The original examiner used field numerical scoring techniques and made a clinical decision from both this and other (e.g. subject's behavior) information.

The panel unanimously made decisions in the same direction in only 22 cases (24%) so these data will not be discussed separately but are included among those 66 cases (72%) where the criterion was a majority of the panel making a decision in the same direction. Among the 66 cases, 100% of the guilty subjects and 35.7% of the innocent subjects were correctly classified (omitting inconclusives) by the original examiner's clinical diagnosis. With the same criterion, original examiner's numerical score correctly classified 100% of the guilty subjects and

27.3% of the innocent subjects (omitting inconclusives). Numerical scoring by an independent examiner correctly classified 97.5% of the guilty subjects and 45.5% of the innocent subjects.

When judicial outcome was the criterion, 100% of the guilty subjects were correctly classified by the original examiner's clinical diagnosis, by the original examiner's numerical analysis, and by the independent evaluator (omitting inconclusives). The original examiner's clinical diagnosis and numerical analysis correctly classified 14% and 17% of the innocent subjects, respectively (omitting inconclusives) with the judicial outcome criterion. The independent examiner correctly classified 43% of the innocent subjects.

Thus, the Bersh, and Barland and Raskin studies both find high agreement between panel decision of guilty and a polygraph examiner decision of guilt, but differ dramatically in agreement between panel and examiner for innocent cases. Several possible reasons have been suggested for this discrepancy (see Lykken, 1979; 1981; Barland, 1982; Raskin, 1978). In the Bersh study, examiner decisions were not totally independent of panel decisions since the examiner had access to limited case data - enough to conduct the polygraph test. This may have increased the correspondence between examiner decision and panel decision beyond that which is representative. However, Barland and Raskin note that the panel in their study often received case files with only small amounts of information. The panel did not necessarily have the police reports, but did have the subject's explanation of the events (Barland, 1982). This lack of information is reflected by the few cases of unanimous panel decision. Barland and Raskin also note that 34 of the cases were tested with a polygraph in poor mechanical condition suggesting that the polygraph examiner had poor charts in certain instances. Thus, the figures reported by Barland and Raskin may underrate the typical agreement between examiner results and a panel decision.

Panel judgments pose a further problem as a criterion. Although the panels

were to judge cases only on merit and ignore legal technicalities, our judicial system classifies a person as guilty only with overwhelming evidence. Thus, panels composed of persons with legal training may be biased towards calling people innocent if there is any doubt about their guilt.

There is only one study where actual guilt/innocence of the subject is independent of the polygraph examination and in which there is certain knowledge of ground truth (Ginton, et al., 1982). This study has been classified as a field study because it used a real-life situation and subjects thought there would be consequences from failure to pass the polygraph test.

Twenty-one Israeli policemen were administered a series of pencil-and-paper tests presented as part of a police course. Beneath the answer sheet for one of the tests was a hidden chemical page that recorded what was written on the answer sheet. The chemical page was removed, the answer sheet returned and subjects scored their own tests. Seven subjects actually changed their answers during the scoring. Several days later subjects were informed that cheating was suspected and were offered an opportunity to take a polygraph test. They were told their career may depend on test results. Fifteen subjects, 13 innocent and 2 guilty, were actually tested with control-question procedures administered at police laboratories by professional polygraphers. Electrodermal activity, cardiovascular activity, and respiration were recorded. There were three participants in each case - an examiner, an observer in an adjacent room, and a blind chart evaluator. Thus, each participant had different information by which to make a decision of truth or deception. The examiner made his decision based on the polygraph charts and subject behavior, the observer made his decision based only on subject behavior, and the chart analyst used only the polygraph charts. In addition to the their overall impressions, the examiner and the blind chart evaluator also scored the tests using field scoring techniques. Each relevant-control pair was scored from +3 to -3 and summed across physiological response systems, questions, and

charts. Total scores from +5 to -5 were called inconclusive. Additionally, 8 other examiners analyzed the charts using both general chart scrutiny, and the numerical scoring method.

Only the data from the examiner, the blind chart evaluator, and the 8 additional examiners will be presented. Using only his clinical judgment, the original examiner correctly classified 100% of the guilty subjects and 85% of the innocent subjects; there were no inconclusive decisions. Numerical field scoring increased the number of inconclusives; one guilty subject and six innocent subjects were classified inconclusive. Omitting inconclusives, the original examiner correctly classified 100% of the guilty subjects and 86% of the innocent subjects using numerical field scoring techniques. The blind evaluator correctly classified 50% of the guilty subjects (none of the guilty were classified inconclusive) and 70% of the innocent subjects using only general chart scrutiny (omitting inconclusives). Numerical field scoring increased the number of inconclusive decisions and improved the accuracy (omitting inconclusives) of the blind evaluator: 100% of the guilty subjects and 83% of the innocent subjects were correctly classified.

The 8 additional examiners correctly classified 94% of the guilty subjects with chart scrutiny and 100% (omitting inconclusives) of the guilty subjects with numerical analysis. Omitting inconclusives, their decisions from general chart scrutiny correctly classified 82% and their decisions from numerical analysis correctly classified 83% of the innocent subjects.

There are a number of additional field studies but they have severe methodological defects. Ben-Ishai (1962) reports 94% correspondence between examiner judgment and case outcome as determined by a student who had collected additional data on the case, but presents no detail on methods used in the polygraph examination or in case verification. Bitterman and Marcuse (1947) report

100% agreement between polygraph decisions and actual truth with 81 subjects in an actual crime but this study contained no subjects guilty of the crime. Elaad and Schahar (1978) report that polygraph examiners correctly classified as deceptive 96.6% of the subjects who were actually guilty; however, this study contained no innocent subjects. It is difficult to draw conclusions from studies using only guilty or only innocent subjects because the examiner would have high correct classification rates merely by a bias to call all subjects innocent or all subjects guilty. There also are older studies reporting high correct classification rates (e.g. Mac Nitt, 1942), but these studies give insufficient detail about procedures and are impossible to evaluate.

Field studies of criterion validity are fairly consistent in their results for guilty subjects. When original examiner subjective decision is used to predict the criteria, 90% to 100% of the guilty subjects have been correctly classified.

Results are similar for studies using a criterion that is independent of examiner decision and studies using a criterion correlated with examiner decision.

Numerical field scoring by the original examiner, chart scrutiny by a blind examiner, and numerical analysis by a blind examiner does not appear to alter the percentage of guilty subjects correctly classified when inconclusives are omitted.

The results for innocent subjects are much more variable. When original examiner's clinical judgment is used to predict the criterion, omitting inconclusives, from 14% to 100% of the innocent subjects are correctly classified. The Barland and Raskin (1976) was the only one to show poor results; problems with this study have been discussed. Omitting the Barland and Raskin data, 85% to 100% of the innocent subjects were correctly classified. It is important to note that the study reporting a 85% correct classification figure used a criterion that was both independent of original examiner decision and accurate (Ginton et al., 1982).

Percentage of correct classification for innocent subjects does not appear to

change when the original examiner uses numerical scoring and when a blind evaluator makes either clinical or numerical decisions. Additional data on the relationship between different types of examiner decision is discussed in a subsequent section.

Relationship Between the Decisions of Blind Examiners and Ground Truth Using Verified Polygraph Cases

The studies discussed below investigate the ability of examiners working in the blind to correctly classify subjects as truthful or deceptive using verified cases. Since ground truth is known, these studies are cited by some researchers as providing evidence on polygraph validity. For example, Lykken (1978; 1981) cites Horvath (1977) as a validity study and uses it as evidence that polygraph tests misclassify large numbers of people.

There are several problems with these studies. First, they do not provide information about the validity of the original examiner's decisions unless the original examiner makes his decision using only information from the polygraph charts (this is more likely when numerical scoring is used). In most cases the original examiner uses both the polygraph charts and additional information in his decision (e.g. subject behavior and case information). Since the two decisions were based on different types of information, these studies will only provide information on the validity of the blind evaluator's decision based solely on the polygraph charts (Barland, 1982).

It is debatable whether these studies are even validity studies of the blind evaluator's decision because the description of sample selection in most of the studies suggests only verified cases confirming the original examiner's decision were used. If cases where the original examiner's decision is incorrect are eliminated from the sample, the sample has a very restricted range. A true validity study also would investigate how the blind examiner scores the cases not correctly classified by the original examiner. However, incorrect decisions were

not necessarily eliminated. All of the verified cases Horvath (1977) found agreed with the original examiner's decision. Thus, there is enough confusion in the reporting of sample selection in all of the studies using verified cases to make them problematical.

It could be said that use of a restricted sample of only those verified cases confirming the original examiner's decision would overestimate the validity of the blind examiner's decision and that validity could not be any higher than what is reported in these studies. However, it is possible that the type of decision the original examiner makes influences the polygraph test. For example, the original examiner making a clinical judgment from subject behavior, case facts, and polygraph charts may make a decision sooner, possibly with fewer charts, than an examiner using only numerical scoring. If the type of decision an examiner makes influences the examination procedure these studies should not be considered as evidence for the validity of the blind examiner's decision.

A set of studies were conducted by John E. Reid and Associates, a private firm specializing in polygraph examination (Horvath and Reid, 1971; Hunter and Ash, 1973; Wicklander and Hunter, 1975; Sawick and Buckley, 1975). These studies use similar basic procedures. Polygraph charts from equal numbers of truthful and deceptive cases verified by confession were selected from the files. The cases contain similar numbers of verified truthful and verified deceptive questions. Polygraph examiners of varying experience administered the standard Reid Control-Question Technique and the examiner made a clinical decision of subject guilt or innocence based on the polygraph charts, subject's overt behavior, and knowledge of case facts. Electrodermal activity, cardiovascular activity and respiration were recorded. These charts were then reanalyzed by examiners blind to subject guilt or innocence, and a decision based only on the information in the polygraph charts was made. Thus, the Reid studies indicate the extent to which the original examiner bases a decision on the polygraph charts as opposed to other

information. It is important to consider that these studies were conducted at John E. Reid & Associates because the Reid technique emphasizes the use of nonverbal indicators (Mullenix and Reid, 1980; Reid, 1980). None of the Reid charts were scored numerically. Most other schools and techniques give little or no weight to nonverbal behavior.

Horvath and Reid (1971) reported the reanalysis of 40 cases by 10 polygraph examiners. Seven examiners had more than one year experience and 3 were interns, having 4 to 6 months of experience. The 40 cases contained no charts that were obvious to any examiner and none that were uninterpretable. Examiners could call individual questions inconclusive but had to make a guilt/innocence judgment for each case. The 10 examiners correctly classified 85% of the guilty cases and 90.5% of the innocent caes. Omitting inconclusives, 89.6% of the truthful questions and 86.7% of the deceptive questions were correctly classified. The experienced examiners were more accurate, averaging 91.4% correct case classifications and the inexperienced examiners averaging 79.2%. Experienced examiners also had higher correct classification rates for individual questions than inexperienced examiners.

Hunter and Ash (1973) reported on data on the blind rescoring of polygraph records. Seven examiners (6 with more than 1 year of experience and 1 intern) reanalyzed data from 20 cases on two occasions separated by at least 3 months. Examiners did not know they were rescoring the same charts. Similar correct classification rates were found on each occasion. Occasion-to-occasion agreement for the same examiner on the individual cases ranged from 75% to 90% (average 85%) and on individual questions ranged from 67% to 91% (average 81%). Combining occasions and omitting inconclusives, examiners correctly classified 86% of the truthful cases, 88% of the deceptive cases, 85% of the truthful questions, and 88% of the deceptive questions. The authors note that the poorest results were obtained from the examiner with the least experience.

The effect of receiving all of the case information on the scoring of polygraph charts was reported by Wicklander and Hunter (1975). Six experienced examiners (mean experience = 4 years) rescored 20 cases using only the polygraph charts. Omitting inconclusives, 95% of truthful cases, 92% of deceptive cases, 91% of truthful questions, and 91% of the deceptive questions were correctly classified. For the second review, examiners were given the polygraph charts, a description of the case, verbal and nonverbal behavioral symptoms noted by the original examiner, and relevant test questions. This additional information primarily reduced the number of inconclusive decisions. Omitting inconclusives, 91% of the truthful cases, 98% of the deceptive cases, 92% of the truthful questions, and 97% of the deceptive questions were correctly classified.

Slowick and Buckley, (1975) investigated the effect of using all physiological measures versus each measure independently on percentage of correct decisions. Seven experienced examiners (mean experience = 3.8 years) first rescored the charts from 30 cases using all physiological data and then analyzed each physiological measure independently (at 3-month intervals to avoid recall). Using all physiological measures and omitting inconclusive results, the seven examiners correctly classified 93% of the truthful and 85% of the untruthful cases, and 93% of the truthful and 80% of the untruthful questions. Analysis of individual physiological measures revealed slightly lower correct classification rates; differences between percentage of correct classifications with each measure were small.

The remaining studies used cases from a variety of sources. Kleinmuntz and Szucko (1982a; 1982b) report the lowest figures of all the studies. Polygraph examinations from 50 verified innocent and 50 verified guilty subjects suspected in crimes of theft were obtained. Six polygraph trainees, in the last few weeks of their internship training program, reanalyzed each chart on an 8-point scale with 1 representing definitely truthful and 8 definitely untruthful. Trainees knew the

proportion of guilty subjects but did not know who was guilty or innocent. When a single question sequence from one polygraph chart was scored, correlations between trainees' ratings and actual guilt or innocence ranged from .26 to .52. Three of the examiners also scored complete charts. Correlations between their ratings and actual guilt/innocence ranged from .45 to .55.

Horvath (1977) reported data on the effect of several variables on analysis of polygraph recordings. Ten field-trained polygraph examiners from law enforcement agencies (5 having more than 3 years experience and 5 having less than 3 years experience, including some interns) judged a sample of polygraph records from 112 criminal suspects from independent cases. The records were sampled so that there were similar numbers of verified and unverified cases, truthful and deceptive subjects, and crimes against people and crimes against property. Data on the unverified cases will be presented in a later section. Polygraph examinations used standard control-question procedures; electrodermal activity, respiration, and cardiovascular activity were recorded. All examinations included at least 2 charts of the control-question test and 1 stimulation test; 48 examinations (43%) included additional charts. The original examiner decision was based on all of the information in the polygraph examination; decision of the 10 other examiners was solely based on a scrutiny of polygraph charts.

For verified cases, 51% of the truthful cases were classified as truthful by the polygraph examiner, and 77% of the deceptive cases were classified as deceptive by the examiner. Results were similar for verified cases of person crime and verified cases of property crime.

Experience of the polygraph examiner did not influence accuracy. There were effects for number of charts. Seventy-one percent of the records containing only the basic two relevant charts were correctly classified, whereas 52.5% of the records containing more than the basic battery were correctly classified (verified and unverified cases combined).

Raskin (1976) reports on the rescoring of records from 12 confirmed guilty subjects and 4 confirmed innocent subjects from the Barland and Raskin (1976) study, by 25 field polygraph examiners with different types of training and experience. Eighteen examiners had at least one year of experience, and 13 had formal training in numerical scoring, of whom 7 explicitly used numerical scoring in their reanalysis of the data. Excluding inconclusives, 90.5% of the 400 decisions were correct. It is not possible from the published data to determine the percentage of innocent and guilty subjects correctly classified because Raskin does not report the percentage of inconclusive results for guilty and innocent subjects separately. If it is assumed that inconclusives are evenly divided, 73% of the innocent subjects and 95% of the guilty subjects would have been correctly classified. Examiner experience did not affect accuracy but other variables did. Examiners who attended schools emphasizing numerical scoring had higher accuracy than examiners attending other schools (97.1% versus 86.9%). Examiners explicitly using numerical scoring had higher accuracy (98.9%) than all the remaining examiners (87.9%) and the subsample of examiners trained in, but not explicitly, using numerical scoring (88.5%).

This study also provides some data on the reliability of the inconclusive category. This information is not available in the other studies because they only used cases where the original examiner made a guilt/innocent decision. When the original examiner's decision was based on more than numerical chart analysis, the blind examiner agreed with 85% of cases classified as truthful, 78.9% of the cases classified as deceptive, and 53% classified as inconclusive. When the original examiner's decision was based only on numerical chart analysis, the blind examiner agreed in 91% of the cases classified as truthful, 87% of the cases classified as deceptive, and 71% of the cases classified as inconclusive.

Edel and Jacoby (1975) reported on agreement between the original examiner and blind examiners, and between blind examiners, when both were making the same type of decision. The forty randomly selected cases were from pre-employment screening examinations for a federal agency and contained 25-30 relevant, irrelevant, control, and overall truth questions. Ten experienced examiners (interview experience from 3 months to 10 years) analyzed the charts. Each examiner conducted the original examination in 4 cases and was the blind rater in 8 cases. Thus, each case had one examiner and one blind rater. Examiners and raters were to decide whether a significant physiological response or no significant physiological response occurred to each question in each physiological system recorded (electrodermal activity, cardiovascular activity, and respiration). The analysis by the original examiner was not done at the time of the examination. It was done later when the examiner also was analyzing the charts from the other examiners' cases.

The examiner and the raters agreed on 96% of the decisions, averaged across physiological measures. The blind raters agreed in 94% of the cases when agreements were averaged across physiological measures. The percentages of correct decisions for each physiological measure were similar.

The studies using blind evaluators to analyze polygraph records of verified cases report a range of correct classifications for guilty and innocent subjects. They are difficult to interpret because the method of sample selection is not clear. If, as many studies indicate, the only cases used were those in which case outcome confirmed the original examiner's decision, then the studies only provide information on how a blind evaluator agrees with the original examiner. When original examiner decision is a clinical diagnosis from case facts, subject behavior, and polygraph charts, it is not surprising that agreement would be low. Furthermore, the type of decision an examiner makes may influence how a test is conducted, making chart usage from one type of examination to make another type of decision problematical.

Relationship Between Different Types of Examiner Decisions When Ground Truth is not Known

Studies investigating the relationship between different types of examiner decision when ground truth is not known and the examiners make decisions using the same type of information are reliability studies.

Test reliability refers to whether test results are repeatable; for example, when the test is administered on different occasions or when it is administered by different people are the same results obtained. Test reliability provides some clues as to potential test validity as a test will be able to predict a criterion only to the extent it is reliable. It should be recognized, however, that even if a test is very reliable, it may not be valid.

There are different types of reliability. Critical issues for polygraph examinations are the extent to which different examiners make the same decision of truth or deception from the same polygraph examination, and the extent to which an examiner will draw the same conclusions when the examination is administered to the same subject on different occasions. There is no information on occasion-to-occasion repeatability of polygraph examinations. There is some information on agreement of different examiners when their decision is made from the same type of information.

Barland and Raskin (1976) report data on the agreement between a blind examiner and the original examiner when the original examiner decision was not necessarily correct. Details of this study are already presented in the Validity section. The original examiner had numerically scored the polygraph charts although the decision, in some instances, was based on more than numerical chart analysis. Omitting inconclusives, the blind examiner classified as truthful 100% of the cases classified as truthful by the original examiner, and classified as deceptive 98% of the cases classified as deceptive by the original examiner. When the original examiner decision was based only on numerical chart analysis, the two examiners agreed in 100% of the cases, excluding inconclusives.

The Kleinmuntz and Szucko (1982) study already discussed reported information on the relationship between the ratings of 6 examiner trainees. Trainees analyzed one chart on an 8-point scale with 1 representing definitely truthful and 8 representing definitely untruthful. Correlations between the scores assigned to the records by the different trainees ranged from .24 to .56. Three of the trainees also scored complete charts; correlations between the three trainees' ratings ranged from .45 to .61.

The Horvath (1977) study provided information on the agreement between the original examiner and ten blind evaluators for unverified cases in addition to the verified case data already discussed. Since the blind evaluators used different information than the original examiner, this study cannot be used as evidence for the reliability of original examiner decision. For the unverified cases, the blind evaluator agreed with 52% of the cases the original examiner classified as truthful and 72% of the cases the original examiner classified as deceptive. However, there was considerable difference in the agreement for person and property crimes with more agreement for innocent subjects in cases of property crime and more agreement for guilty subjects in cases of person crime. Horvath also reported significant consistency among the blind evaluators both for verified ($r=.89$) and unverified ($r=.85$) cases.

All but one of the field studies investigating the repeatability of examiner decisions based on the same type of information report high reliability. Unfortunately, all of the data is either on the repeatability of numerical scoring or the repeatability of blind examiner chart scrutiny. There appears to be no data on the relationship between clinical diagnosis of the original examiner and the clinical diagnosis of another examiner when both have equivalent information about the case. The Kleinmuntz and Szucko (1982) study was the only one to repeat low reliability. As previously noted, the examiners in this study were trainees and used only one chart to make their decision.

There are various other reports of the reanalysis of cases that were verified and confirmed the original examiner's decision. None of these studies report how the original examiner made the decision (this was not necessarily the purpose of the study) and it cannot be determined whether errors reflect differences in information available to make a decision or whether they are random. Davidson (1982) provided evidence on the reanalysis of 10 confirmed deceptive and 11 confirmed truthful property crime cases. Seven examiners (experience 6 months to 14 years) numerically scored records (U. S. Army Polygraph School procedures) containing different types of physiological recordings. The examiners first evaluated only Cardio Activity Monitor recordings (a device designed for the U. S. Air Force), then only the electrodermal, respiration, and cardiovascular activity records, and, finally, all recordings to make judgments of truth deception or inconclusive. For guilty cases, 90% of the decisions with the Cardio Activity Monitor, 87% of the decisions with the polygraph and 90% of the decisions with both were correct. For innocent cases, 85.7% of the decisions with the Cardio Activity Monitor, 84% of the decisions with the polygraph, and 88% of the decisions with both were correct. Inconclusive decisions were considered incorrect decisions and it is not possible from the presented data to determine the percentage of guilty and innocent cases correctly classified, omitting inconclusives.

Ben Ishaï (1962) reports 10 criminal cases (number of guilty and innocent not known) were accurately classified by comparing average electrodermal response to relevant and control questions. Widacki (1982) reported that 91.6% of guilty ($n = 16$) and innocent ($n = 22$) confirmed cases using control-question tests were correctly classified when charts were reanalyzed using the Backster numerical scoring method. Separate percentages are not given for guilty and innocent subjects. Guertin and Wilhelm (1954) reported that factor analytically derived measures of electrodermal activity from relevant-irrelevant tests on 19 guilty and 15 not guilty criminal suspects correctly classified 95% of the guilty subjects and 94% of the innocent subjects. .

Laboratory Studies

In laboratory research the criterion (ground truth) is both independent and accurate because it is controlled by the experimenter. Many laboratory studies have used a mock crime procedure. Subjects are designated as guilty or innocent by the experimenter and the guilty subjects participate in the crime (typically stealing money). The consequence for failure to produce a truthful outcome on the test (i.e. guilty subjects deceiving the examiner and innocent subjects demonstrating their innocence) is generally loss of a monetary bonus. Only a few laboratory studies have used a professional polygraph examiner to test subjects and score records. Many of the examinations are conducted by graduate students under the supervision of psychologists or psychophysiologicals with experience in psychophysiological recording techniques. Control-question, peak-of-tension, guilty knowledge, and relevant-irrelevant techniques have been used. Most of the studies measured respiration, cardiovascular activity, and electrodermal activity, although many of the studies using the peak of tension or guilty knowledge test formats measured only electrodermal activity.

Many of the studies have used control-question techniques in the context of determining how other variables such as drugs (Gatchel, Smith and Kaplan 1983), subject characteristics (Raskin and Hare, 1978; Hammond, 1980), information and practice (Rovner, Raskin and Kircher, 1978), and countermeasures (Honts, 1978) effect detection and deception. Although this information is important, it is the purpose of this report to review polygraph accuracy on samples that have not received special procedures. Thus, when studies have manipulated other variables, only information from the normal guilty and normal innocent subjects receiving standard polygraph tests will be analyzed. The results of the manipulated variable are listed in the summary of each study.

Control-Question Test - Criterion Validity

A set of studies by Raskin and his colleagues (Barland and Raskin, 1975; Rovner, Raskin, and Kircher, 1978; Raskin and Hare, 1978; Podlesny and Raskin, 1978) have used similar procedures to investigate detection of deception in a laboratory situation. Subjects are designated guilty or innocent; guilty subjects steal money or an item from a particular location and receive a monetary award for deceiving the examiner. Motivation for innocent subjects to appear innocent varies with the study. The examiners used in the Raskin studies were relatively inexperienced, but Barland, Raskin and Podlesny had attended basic polygraph training courses. Several of the studies recorded many physiological responses, but classified guilty and innocent subjects only from a combination of electrodermal, respiratory, and cardiovascular recordings. All of the studies use a ten-item control question test. Charts are evaluated with numerical field scoring techniques in which each physiological response measure on each control-relevant question pair is scored from +3 to -3 based on the magnitude of the difference, and summed across physiological response measures and questions. Scores below a certain level indicate deception, above a certain level indicate truthfulness. An inconclusive range also is used.

Barland and Raskin (1975) reported on 36 innocent and 36 guilty subjects receiving three presentations of the test. Charts were scored both by the original examiner (experience not discussed) and 5 additional examiners who were professionally trained and experienced in numerical scoring techniques. The authors note that the electrodermal and cardiovascular recordings were not of high quality. Innocent subjects were told the details of the crime and that, if they did not appear innocent on the test, they may not get course credit for participating. An inconclusive range from +4 to -4 was used. The original examiner correctly classified 88.5% of the guilty subjects and 71% of the innocent

subjects when 25 inconclusives were omitted from the analyses. Correct classifications for the five additional examiners ranged from 79% to 86% (mean = 82%, results are not presented separately for innocent and guilty subjects).

Rovner, Raskin, and Kircher (1978) reported on the effects of information and practice on detection of deception. Only the group receiving the standard polygraph test will be discussed in this report. There were 12 innocent and 12 guilty subjects in the standard group and they received three presentations of the test. The inconclusive range in this study was from +5 to -5. Money was used to motivate innocent subjects to appear truthful on the polygraph test. Innocent subjects were not told the details of the crime; they were merely told that a theft had occurred. Classification results are from the scoring of another evaluator who did not observe the test. Omitting the 2 inconclusives, 100% of the guilty subjects and from 90% to 92%* of the innocent subjects were correctly classified.

Similar correct classification rates were reported by Raskin and Hare (1978) in a study on the effect of psychopathy on detection of deception using a professional polygraph examiner. Only the results from the 12 innocent non-psychopaths and 12 guilty non-psychopaths will be discussed. Money was used to motivate innocent subjects to produce a truthful outcome on the test. The authors state that the innocent subjects were informed about the crime. The examiner continually stressed the importance of being truthful to all subjects thereby directing the attention of innocent subjects to the control questions and guilty subjects to the relevant questions. Subjects received from 3 to 7 presentations of the test and were initially classified from the first three charts using inconclusive range of between +5 and -5.

*

Since the study does not report the actual status of the inconclusive subjects, it is not possible to cite a precise figure for the percentage of innocents subjects correctly classified.

If this classification was inconclusive, additional charts were scored. Between each chart, questions were discussed with the subjects and reworded if necessary. Omitting the 4 inconclusives, 100% of the guilty subjects and 91% of the innocent subjects were correctly classified. Comparison of classification accuracy for subjects receiving more than three charts with those only receiving three charts demonstrated that additional charts reduced the number of inconclusive decisions.

Podlesny and Raskin (1978) reported results both on the control-question test and the guilty knowledge test. Only the control-question results will be reported here; guilty knowledge tests result will be discussed in a subsequent section. Twenty subjects were guilty and twenty subjects were innocent. Half of the subjects received a control-question test with nonexclusive control questions. Half received tests with exclusive control-questions. A nonexclusive control-question has no limit as to time, and thus includes the period in which the offense occurred. The exclusive control-question excludes the offense under investigation by referring only to an earlier period in the subject's life. The innocent subjects received a brief description of the crime but they did not know its details. Money was used to motivate innocent subjects. Subjects received 3 or more presentations of the test. Between charts subjects were questioned and the wording of the control-questions was changed if necessary. Results for different methods of quantifying results are reported; those using the conventional inconclusive range of from +5 to -5 will be presented. Charts also were scored by a blind rater who had not observed the polygraph test. Omitting the 4 inconclusives, 82% of the guilty subjects and 95% of the innocent subjects were correctly classified. Slightly higher correct classification rates were found for exclusive control-questions than non-exclusive control-questions.

There are a number of other studies using control-question procedures in a mock crime situation. Dawson (1977) presents a brief report on detection accuracy with

ten guilty and ten innocent subjects. Different methods of scoring the responses were used; the results with the standard procedure are of interest here. When average vasomotor and electrodermal responses to relevant and control questions were compared, 89% of the guilty subjects and 71% of the innocent subjects were correctly classified omitting the 4 inconclusives. An additional study (Dawson, 1977, 1980) investigated the ability of actors using the Stanislavsky method of acting to deceive the polygraph examiner, but will not be discussed since the purpose of the test was to evaluate a countermeasure method and a variation in polygraph technique.

Hammond (1980) reported on normal college students, alcoholics, and psychopaths, but only the data from the college students will be presented. There were 11 innocent and 10 guilty subjects. Instructions for the mock crime were presented by tape and were designed to motivate the subjects. Innocent subjects listened to the same tape and, thus, knew the details of the crime. All subjects received money for a truthful outcome of the test. Inexperienced examiners from the Backster School of Lie Detection in the fifth and sixth weeks of their 7-week training course conducted the tests using zone-comparison procedures and the zone-comparison method of scoring. Respiration, electrodermal activity, and cardiovascular activity were recorded. Testing involved a minimum of three charts, but in most cases only two charts were used for analysis. The inconclusive range in this study was between +8 and -8. The charts also were analyzed by an expert examiner. Omitting the 7 inconclusives, the student examiners correctly classified 86% of the normal guilty subjects and 57% of the normal innocent subjects. The expert examiner had 15 inconclusives and correctly classified none of the normal innocents and 100% of the normal guilty subjects.

Bradley and Janisse (1981) reported on 96 guilty and 96 innocent subjects. Half of the innocent subjects and half of the guilty subjects were told that they

would receive an electric shock if they were judged guilty. Both control-question and guilty knowledge tests were administered in counterbalanced order. Only the data from the control-question will be discussed in this context. The test consisted of three presentations of nine questions; subject were allowed to discuss questions with the examiner in order to eliminate ambiguities. The numerical scoring in this study used a more limited range than other studies; each control-relevant pair was scored 1, 0, or -1 and the inconclusive range was from +2 to -2. Electrodermal activity, heart rate, and pupil size change were recorded and subjects were classified as guilty or innocent for each physiological measure independently. Only the electrodermal results will be presented, since the most correct classifications occurred with this measure, and heart rate alone has seldom been a factor in field testing. Pupillometry is not in common use anywhere in the field. Omitting the 56 inconclusives, 82% of the guilty subjects and 86% of the innocent subjects were correctly classified. The authors note that their procedures for the control-question test were not optimal because they did not combine the physiological measures to classify the subjects, and their semi-objective scoring procedure had a limited range.

The effect of physical countermeasures on the detection of deception was investigated by Honts (1982). Only the data from the 12 innocent and 12 guilty subjects using no countermeasures will be discussed. Innocent subjects knew the general nature of the crime, but were naive to its details. All subjects were told they would receive extra points applied to their final course grades if they produced a truthful outcome on the test. Respiration, electrodermal activity, and cardiovascular activity were recorded by a licensed examiner with 5 years experience. Subjects received four presentations of eight questions. Immediately after the test, the examiner was required to make an innocent/guilty/inconclusive decision based on both subjects' behavior and limited (less than 5 minutes) chart analysis. The examiner also scored the charts by semi-objective methods using an

inconclusive range of from between +12 and -12. Another experienced evaluator scored the charts using the semi-objective method. Results are based on the three most productive charts. The original examiner's clinical decision correctly classified 86% of the guilty subjects and 60% of the innocent subjects, omitting the 7 inconclusives. Semi-objective scoring by the original examiner evaluator correctly classified 100% of the guilty and 67% of the innocent subjects, omitting the 11 inconclusives. Omitting inconclusive decisions, the original examiner and the blind evaluator agreed in 100% of the cases when decisions were made with semi-objective scoring.

The effect of propranolol on detection of deception was studied by Gatchel, Smith, and Kaplan (1983). Only the results from the 7 guilty and 7 innocent subjects not receiving drugs will be discussed. All subjects would receive a bonus for appearing truthful on the test. Innocent subjects were told that a mock crime had taken place but were not told its details. Blood pressure, electrodermal activity, and respiration were recorded by two professional examiners administering a nine-item control-question test. Charts were numerically scored with an inconclusive range of +5 to -5. Subjects appear to have received only one presentation of the test. Omitting the 6 inconclusives, 75% of the guilty subjects and 100% of the innocent subjects were correctly classified.

Kleinmuntz and Szucko (1982) reported on 15 innocent and 15 guilty subjects. Subjects were administered a control-question polygraph test by four examiners who were trainees in the first half of training. Tests were scored by six examiners (3 trainees with three months of experience and 3 with more experience up to 8 years) but only one chart of each set of three or more was used for the analyses. Tests were classified on an 8-point scale with 1 being definitely truthful and 8 being definitely untruthful. When a rating of 5 or more was used to indicate deception and less than 5 indicated truthfulness, 71% of the guilty subjects and 51% of the innocent subjects were correctly classified. No inconclusive range was used.

Correlations between examiner ratings and actual guilt/innocence ranged from .02 to .43. Experience of the examiners scoring the tests did not influence the results.

Waid, Orne, and Wilson (1979) reported results on 15 innocent and 15 guilty subjects. This study did not use a mock crime. Rather, guilty subjects learned six code words during timed interpolated tasks. Innocent subjects performed the same tasks but learned no words. Innocent subjects were told it was often difficult to prove ones innocence on a polygraph test and, the authors note, care was taken to involve the innocent subjects in the task. Guilty subjects were told that highly mature and intelligent people could deceive the examiner. A professional polygraph examiner administered several types of polygraph tests. The first test consisted of three relevant and two control questions presented twice. Between presentations, subjects were questioned about their answers and received a stimulus test. Subjects were classified as deceptive if the response to any of the three critical question was larger than the response to the control questions; thus, this study used no inconclusive range. Only skin conductance was used to classify the subjects. Eighty percent of the innocent and 80% of the guilty subjects were correctly classified on each test presentation.

Waid, Orne, and Orne (1981) used a similar situation to investigate detection of deception with 34 innocent and 40 guilty subjects. The polygraph test in this study consisted of 3 presentations of 9 relevant questions, 8 control questions and 5 irrelevant questions and was administered by a professional polygraph examiner. Different physiological measures were recorded and classification of guilty or innocent was done for individual physiological measures. The best results were obtained for skin conductance. Subjects were classified as deceptive if the average electrodermal response to the relevant questions was larger than the average electrodermal response to the control questions and were classified as innocent if the opposite relationship occurred; thus, no inconclusive range was used in this study. Seventy-three percent of the guilty subjects and 76.5% of the

innocent subjects were correctly classified. This study also administered guilty knowledge tests after the control-question procedure; these results will be discussed later.

Widacki and Horvath (1982) compared the polygraph with fingerprints, hand-writing specimens, and eyewitnesses identifying persons from full-faced photographs in solving 20 mock-crimes. Subjects were divided into 20 groups of four people. In each group, three subjects were innocent and one subject was guilty. Each guilty subject was involved in an independent but similar case. The guilty subject took an envelope to a specific building, gave it to a specific person, and received a small parcel containing a cosmetic. Upon receiving the parcel, he signed the receipt with a fictitious name. Polygraph examiners made subjective decisions of guilty or innocence from a polygraph test using Reid control-question procedures and recording respiration, cardiovascular activity, and electrodermal activity. The examiner tested each group of four before making a decision about who was guilty. Using the polygraph, 18 of the 20 cases were solved (there was one error and one inconclusive decision). Thus, omitting inconclusives, 95% of the cases were solved. Handwriting and fingerprinting also were accurate although there were many inconclusive decisions with fingerprinting. The eyewitness identification results were very poor.

Barland (1981) is the only study to use control-question test in a mock screening situation. Polygraph tests were administered by three experienced examiners and respiration, electrodermal activity and cardiovascular activity were recorded.

Subjects initially supplied biographical information which was verified by a background investigation. Thirty subjects then supplied false biographical information and 26 subjects supplied accurate biographical information to the polygraph examiner. The polygraph test consisted of a minimum of three charts of thirteen questions. If no decision could be made, additional charts were run with

a maximum of six charts. The examiner was not allowed to interrogate the subject. Three methods were used to evaluate the charts - the zone method, the greatest control method, and a relevant-irrelevant method. The data for the relevant-irrelevant method will be presented in a subsequent section. The zone method is similar to the method used in the previously cited control question laboratory studies. Each control-relevant pair was scored from +3 and -3 and summed; scores between +3 and -3 were inconclusive. The greatest control method used the same procedure except that the relevant questions were evaluated against the one control question with the largest reaction.

When subjects were classified deceptive or non-deceptive (irrespective of whether the specific question was identified), omitting the 9 inconclusives, 81% of the guilty and 76% of the innocent were correctly classified with the zone method. With the greatest control method, omitting the 10 inconclusives, 68% of the guilty and 83% of the innocent were correctly classified. There were 250 truthful questions and 30 deceptive questions (each deceptive subject lied to 1 question and answered the remaining questions truthfully). Omitting inconclusives, the zone method correctly identified 91% of the nondeceptive questions and 63% of the deceptive questions, whereas the greatest control method correctly identified 54% of the deceptive questions and 97% of the nondeceptive questions.

The laboratory studies of control-question techniques report a range of correct classification figures for guilty and innocent subjects. Most studies report that from 85% to 100% of the guilty subjects are correctly classified. Results for innocent subjects are much more variable. Few of the studies used professional examiners, but it must be recognized that most of the other studies used graduate students under the supervision of psychophysiologicalists who were well-versed in psychophysiological recording techniques. The lowest correct classification rates for innocent subjects are reported by two studies using student examiners (Kleinmuntz and Szucko, 1982; Hammond, 1980). The Kleinmuntz and Szucko study also

classified subjects from review of only one polygraph chart out of each set of three or more. Other studies (Waid et al., 1979; 1981; Bradley and Janisse, 1982) classified subjects as guilty or innocent by using each physiological measure independently. It is well-known that all people do not respond in the same physiological system (Engel, 1960; Garwood, Engel, and Capriotti, 1982), and higher correct classification rates would likely be found if subjects were classified from several psychophysiological measures as is done in field studies. For example, in a Japanese study (Ohnishi, Matsuno, Arasuna, and Suzuki, 1976), the detection rate for breathing was 46% and for electrodermal was 72%. When combined, the detection rate rose to 92%. The possible importance of another variable should be considered. In most of the studies, innocent subjects knew that a crime had been committed but were naive to its details. In the Hammond, (1980) study, however, innocent subjects heard the same tape that guilty subjects heard and thus, knew all the details of the crime. The subjects in the Barland and Raskin (1975) study also knew the details of the crime and this study reported the lowest correct classification figures of the studies by Raskin and colleagues. However, it appears that subjects in the Raskin and Hare (1978) study also knew the details of the crime (although details are minimal) and this study reported high correct classification figures. This study also used more sophisticated procedures than the earlier Raskin studies. The effect of knowledge of crime details by innocent subjects on classification accuracy certainly deserves further consideration.

Several interpretations of these laboratory studies of control question technique studies are possible. In the worst case, the range of accuracy in supporting truthfulness is 0 to 100%, and the range of detecting deception is 71% to 100%. If those cases conducted by student examiners are removed from consideration then, the range for truthfulness improves to 71% to 100%, and the range for detecting deception is 73% to 100%. If the weighted average for all of the studies is used, the truthful rate is 77.1% and the deceptive rate is 81.3%.

If the cases done by students are removed; the truthful rate is 81.2% and the deceptive rate is 81.9%. However, weighted averages presume a similarity in the research methods, which was not always the case and these figures should not be cited as reflecting the percentage of guilty and innocent people the polygraph correctly classifies. One conclusion is clear, the closer the research approximates field conditions by using field techniques, field instruments, trained field examiners, and the standard field methods for chart analysis, the higher the accuracy.

Control-Question Test - Reliability

A few of the studies investigated the relationship between the decisions of the original examiner and those of an independent evaluator. All of these studies used numerical field scoring techniques and examiner decisions were not based on subject behavior. The procedures of each of the studies investigating reliability were discussed in the previous section.

All of the studies demonstrate that numerical field scoring techniques are very reliable. Barland and Raskin (1975) reported that the correlations for the numerical scores of pairings by examiners (one original and five additional) ranged from .78 to .95 (mean = .86). Of the 1080 pairwise comparisons, examiners agreed on 98% of the cases when inconclusive decisions were eliminated. Podlesny and Raskin (1978) reported a correlation of .97 between the numerical scores of the original examiner and one additional examiner, who had not observed the test. When both examiners had made decisions, there was agreement in 100% of the cases. Honts (1982) reported that the numerical scores of the original examiner and a blind evaluator had a correlation of .88 and that there was 100% agreement when both examiners had made decisions. Hammond reports that the correlation for student examiners' numerical scores and expert examiner's numerical scores were correlated .64 for the innocent condition and .61 for the guilty condition. It should be

noted that these correlations are computed for the guilty and innocent conditions separately. If the correlation between the two examiners' numerical scores was computed across all subjects, the correlation would be very high.

Relevant-Irrelevant Technique - Criterion Validity

Only a few laboratory studies have used relevant-irrelevant techniques in the detection of deception. The chapter on Utility contains a description of one of these techniques. The studies vary considerably in their procedures.

Blum and Osterloh (1968) presented data on the ability of the relevant-irrelevant technique to detect false information in stories involving crimes in a laboratory situation. Subjects were police informants who had previously given information on criminal cases to local or federal agencies. It should be recognized that when informants are talking to the police their lives may be in danger; thus, subject arousal in this study may have been high. The situation was structured so that some subjects told completely true stories (n = 9), some told partially false stories (n = 7), and some told completely false stories (n = 4). True stories were those that had been told on a previous occasion and had withstood the scrutiny of investigation. Partially false stories were true stories that had been told in the past, but with some items changed. Totally false stories were invented, but contained credible information.

Three professional polygraph examiners conducted the tests. The authors note that each story provided from five to eight relevant questions, but do not give details of the polygraph test other than it took two hours. Physiological measures used are not mentioned; however, since the examiners used a three-channel Stoelting polygraph, presumably electrodermal activity, respiration, and cardiovascular activity were recorded. Examiner decision was a clinical judgment of truth or deception based on a scrutiny of the charts.

The examiners were able to correctly categorize informants telling totally true, totally false, and partially false stories in 100% of the cases. There was some error in classifying the 42 deceptive questions and 64 truthful questions. Ninety-three percent of the deceptive questions and 98% of the truthful questions were correctly classified.

Similar results were obtained by Correa and Adams (1981) who used the relevant-irrelevant technique in a mock pre-employment screening session. Subjects initially completed a Preemployment Data Sheet and nine items were selected. On the subsequent polygraph test, the 20 designated deceptive subjects lied regarding these items and the 20 designated truthful subjects answered questions about them truthfully. Motivation also was systematically varied with certain subjects competing for a \$25 bonus based on their performance. The examiner's training and experience was not mentioned. The polygraph test consisted of three series, each series containing 32 questions about the Preemployment Data Sheet with 3 of the potential lie questions in each series. Subjects were questioned between charts about any questions where the response to the relevant question was larger than the response to the irrelevant questions. These questions were rephrased and inserted at the beginning of the next series. Electrodermal activity, respiration and cardiovascular activity were recorded, and examiner decision was a subjective judgment of truth or deception.

All of the subjects were correctly classified as lying or telling the truth. Accuracy was not as high for individual lies. Eighty-eight percent of the lies in the lying motivated and 79% of the lies in the lying unmotivated were correctly identified.

The Barland (1981) study discussed earlier, used a relevant-irrelevant evaluation procedure in addition to two numerical scoring methods to detect false biographical information. It should be noted that the polygraph test was a control-question test format, not relevant-irrelevant. There were 30 deceptive and 26 truthful subjects with each guilty subject responding deceptively to one of five questions. Contrary to Blum and Osterloh (1968), and Correa and Adams (1981), the examiner was not allowed to question subjects between charts. Omitting inconclusives, 86% of the deceptive and 76% of the truthful subjects were correctly identified, as compared to 68% and 83% respectively, for the zone comparison

numerical evaluation. Individual question analysis revealed that, omitting inconclusives, 69% of the deceptive and 92% of the truthful questions were correctly identified. Of the three methods of chart analysis, the relevant-irrelevant method was the only one able to correctly detect both truthful and deceptive answers to individual questions. The fact that the relevant-irrelevant method of chart analysis was the last one to be applied may have contributed to this finding.

Heckel, Brokaw, Salzberg and Wiggins (1962) used the relevant-irrelevant technique with subjects who believed they were accused of a real theft. The study used normals, and delusional and non-delusional psychotics, but only the results from normals will be discussed. Subjects were administered a series of psychological tests during which the experimenter briefly left the room leaving his wallet on the desk. When the experimenter returned he accused the subject of taking \$20 from the wallet, although nothing was actually missing. A professional polygraph examiner administered two series of an 11-item General Question Test (a type of relevant-irrelevant) with an interspaced stimulation test. Respiration, heart rate, and electrodermal activity were recorded. Charts were scored by four "experts in lie detection" who did not know the nature of the study. Each examiner made a clinical judgment of deception, no deception, or inconclusive. All of the five normals were correctly classified as non-deceptive by the four judges.

Thus, two of the studies of the relevant-irrelevant technique report that they were able to correctly classify 100% of the lying and 100% of the truth telling subjects. Correct classifications for specific truthful and deceptive questions were slightly lower. Heckel et al. (1962), using only innocent subjects, also classified all subjects correctly. One factor that may be important is subject interrogation between the polygraph charts. Barland (1981) did not allow interrogation. Questioning subjects and restructuring questions, as is typical in field usage, would be expected to increase polygraph accuracy.

Relevant-Irrelevant Test - Reliability

Heckel et. al (1962) appears to be the only laboratory study with information on the reliability of the relevant-irrelevant technique, but they used only innocent subjects. There was 100% agreement between the four examiners. Edel and Jacoby (1975) report a field study on reliability of the relevant-irrelevant test in which they found an agreement rate of 96% between the original examiner and independent evaluators and 94% between the independent evaluators. A detailed discussion of this research appears in an earlier section.

Guilty Information Tests - Criterion Validity

Both the guilty knowledge test and the peak of tension tests use information about a crime or situation that only the guilty person would know. The techniques are described in the Chapter on Utility. The critical difference between the tests is item placement. In the guilty knowledge test the guilty information is randomly placed among the alternatives. In the peak of tension test the item is placed in a specified sequence known by the suspect.

All of the studies investigating the guilty knowledge test have been laboratory studies. Professional polygraph examiners have rarely been used and electrodermal activity frequently is the only physiological response recorded. Several studies have used the specific guilty knowledge test procedure of Lykken (1959). Several multiple choice questions are administered. Each question contains one critical alternative (information known by guilty but not by innocent people) and several similar, but not relevant, alternatives. The position of the relevant alternative to each question is randomly determined. Each question is scored, 2, 1, or 0 depending on the size of the response to the critical information in relation to the size of the response to the other items. The first alternative is not scored to reduce the effect of orienting. A question is scored 2 when the largest response occurs to the critical alternative and a 1 when the second largest response occurs to the critical alternative. All other possibilities are scored 0. Scores are summed across questions with those above a certain level indicating deception and those below a certain level indicating truthfulness. There is no inconclusive range.

Lykken (1959) reported on subjects guilty of two mock crimes, a murder and a theft (n = 13), subjects guilty of only one crime, either the murder (n = 12) or the theft (n = 12), and subjects innocent of both crimes (n = 12). Subjects were told they would receive an electric shock if any of their responses indicated deception but really received random shocks. The guilty knowledge test consisted of two lists of six multiple choice questions, one for the murder and one for the

theft. The number of alternatives to each question averaged 4.67 in the murder list and 5 in the theft list. A score over 6 indicated deception and a score of 6 or less indicated innocence. Subjects made no response to the alternatives. These procedures correctly classified 100% of the innocent, 100% of the subjects committing only the theft, 92% of the subjects guilty of murder, and 69% of the subjects involved in both crimes. Thus, the only errors involved calling the guilty innocent. Lykken (1960) obtained 100% detection when he used the guilty knowledge test to match subjects with their biographical information in an experiment where subjects were taught how to defeat the test. Results will not be discussed since the study used a special sample.

Davidson (1968) reported high correct classification of innocent and guilty subjects in a mock crime situation with a guilty knowledge test consisting of 6 multiple-choice questions each with five alternatives. Apparently the questions were presented only once. Forty-eight students were divided into groups of four with each group containing three thieves and one innocent subject. Each group participated in a theft of money from a different victim with the thieves in each group personally planning the details of the crime. The situation for each group was structured so that one thief tried but failed, one thief had no opportunity to attempt the crime, and one committed the crime. Only the thieves that completed the crime were considered guilty so there were 12 guilty and 36 innocent subjects. Motivation also was varied with the group of guilty subjects receiving variable amounts of money for deceiving the examiner. Examiner decision was based on electrodermal activity scored by the method devised by Lykken (1959) with a score of 6 or less indicating innocence and greater than 6 indicating guilt. Subjects made no response to the questions. Ninety-two percent of the guilty subjects and 100% of the innocent subjects were correctly classified.

Podlesny and Raskin (1978) used the guilty knowledge test with 10 innocent and 10 guilty subjects in a mock crime situation. The test consisted of one

presentation of 5 questions, each with 6 alternatives. The test was scored using the Lykken (1959) procedure and electrodermal activity was the physiological measure. Scores of 6 or greater indicated deception and less than 6 indicated truthfulness. This study also investigated the effect of an inconclusive zone on the guilty knowledge test with scores of 5 and 6 considered inconclusive. Subjects were required to answer no to each alternative. One hundred percent of the innocent and 80% of the guilty were correctly classified when no inconclusives zone was used. When the inconclusive zone was used there was similar accuracy and 10% were inconclusive.

A real cheating situation was used by Balloun and Holmes (1979) to investigate the guilty knowledge test in the detection of deception. Subjects were completing a bogus intelligence information test along with two confederates as part of a laboratory situation. Students were encouraged to do well on the test and told if their performance was poor they would have to meet with a panel of psychologists. One of the confederates urged the subjects to cheat by supplying a copy of the answer sheet. Nearly half of the subjects actually cheated; thus there were 16 non-cheaters and 18 cheaters. The guilty knowledge test consisted of two series of questions each series containing five relevant questions each question with five alternatives. The relevant questions contained only the information the guilty subjects could have known from the answer sheet provided by the confederate. Subjects were required to verbally respond after each alternative by repeating the alternative. The Lykken (1959) scoring system was used with scores greater than 5.5 indicating cheating and less than 5.5 indicating no cheating. When classification was based on electrodermal activity, 61% of the cheaters and 87.5% of the non-cheaters were correctly classified on series 1. On series 2, 17% of the cheaters and 94% of the non-cheaters were correctly identified. Thus, there was a significant reduction in the detection of cheaters from series 1 to series 2. Heart rate and finger pulse volume were not effective in detecting cheating.

The Bradley and Janisse (1981) study used a guilty knowledge test in addition to the control-question procedures. The control-question and guilty knowledge tests were counterbalanced. There were 96 innocent and 96 guilty subjects of a mock crime. The guilty knowledge test consisted of one presentation of 4 questions, each question containing 5 alternatives. It is not clear whether subjects were required to make a response. The scoring method of Lykken (1959) was used. A score of 4 or more indicated deception and less than 4 indicated innocence. Skin resistance was the best discriminator with 89% of the innocent subjects and 59% of the guilty subjects correctly classified. Pupil size and heart rate change were not as accurate in determining guilt and innocence. The authors noted that their version of the guilty knowledge test was a weak version since it did not contain many items of information and was only presented to the subjects once.

Stern, Breen, Watanabe, and Perry (1981) had 26 innocent and 26 guilty subjects play the parts of American Embassy hostages in Iran. Guilty subjects received written instructions about their involvement in a murder plot. Innocent subjects received instructions that had some of the same relevant words as the instructions for guilty subjects, but involved no murder plot. Feedback also was manipulated but these results will not be discussed. The guilty knowledge test consisted of one presentation of 10 questions, each with 5 alternatives. Four of the ten questions were control questions. Subjects were to make no response during test presentations and responses were scored by the method of Lykken (1959). Ninety-two percent of the guilty and 85% of the innocent subjects were correctly classified.

Several studies by Waid and associates (Waid, Orne, Cook, and Orne, 1978; 1981; Waid and Orne, 1982; Waid, Orne and Wilson, 1979) used similar procedures to investigate the effect of several variables on the guilty knowledge test. Only the results for classification accuracy will be reported. Guilty subjects learned six

code words during a series of tasks. Depending on the study, innocent subjects either completed a questionnaire or performed the same tasks as the guilty subjects, but learned no code words. Guilty subjects were told that highly intelligent mature individuals could deceive the examiner. Care was taken to involve innocent subjects in the task. The guilty knowledge test consisted of a variable number of presentations of a list of 24 words. There were six categories of words with each category containing four words, one word being the critical code word. A word was considered detected if the code word elicited a larger response than the three alternatives. Subjects were required to make a response in all of the studies. Waid et al. (1978) reported data from three experiments all using electrodermal activity to detect deception. Experiment 1 used 29 guilty subjects and 11 innocent subjects with two presentations of the 24 word list. Electrodermal activity was measured. Seventy-nine percent of the guilty and 73% of the innocent were correctly classified. Experiment 3 used the same procedures except that 15 innocent and 15 guilty subjects were tested and each of the 6 categories of words contained 6 words. Seventy-three percent of the guilty and 80% of the innocent were correctly classified. The 18 guilty and 10 innocent subjects in Experiment 2 were tested in groups of 3 to 7 and subjects were told they would receive a shock in the end of the test proportional in strength to the number of detected lies. The list of 24 words was presented five times. Sixty-one percent of the guilty and 90% of the innocent subjects were detected. Waid et al (1981) correctly classified 100% of the innocent subjects and 82% of the guilty subjects when electrodermal activity was the physiological measure. Cardiovascular activity and respiration were not as good at classifying subjects. This study used two presentations of the 24-item list. There were 11 innocent and 11 guilty subjects.

Waid, Orne, and Wilson (1979) correctly classified 93% of the 15 innocent subjects and 53% of the 15 guilty subjects using electrodermal activity. The list

of 24 words was repeated four times. The guilty knowledge test was administered by a professional examiner after several types of polygraph tests were administered. Waid and Orne (1980) reported data from two experiments, both recording electrodermal activity. Experiment 1 used the same procedure as Experiment 2 in Waid et al. (1978). There were 18 guilty and 10 innocent subjects. Sixty-seven percent of the guilty and 90 percent of the innocent subjects were correctly classified. The 15 guilty and 15 innocent subjects in Experiment 2 were individually tested by a professional examiner and the guilty knowledge tests were preceded by other types of polygraph tests. The 24-item list was presented four times. Sixty percent of the guilty subjects and 100% of the innocent subjects were correctly classified.

Giesen and Rollison (1980) reported on college students who were selected because they reported showing high palmar sweating responses to stress. There were 20 guilty and 20 innocent subjects. Guilty subjects read and followed detailed instructions involving a murder. Innocent subjects read and followed instructions that contained some of the same information but involved no murder. The common items were used as relevant items on the subsequent guilty knowledge test. The test included 8 questions (6 relevants) each with five alternatives. Electrodermal activity was recorded and the test was scored using the procedure of Lykken (1959). Ninety-five percent of the guilty and 100% of the innocent subjects were correctly classified.

Apparently, there are few studies using the peak of tension test with both innocent and guilty people. The Waid, Orne, and Wilson (1979) study, already discussed used a peak of tension test in addition to two control-question tests and a guilty knowledge test. The peak of tension was always presented after the control question tests. Forty-seven percent of the 15 guilty subjects and 80% of the 15 innocent subjects were correctly classified. On both control-question

tests, 80% of the guilty and 80% of the innocent were correctly classified, whereas with the guilty knowledge test 53% of the guilty and 93% of the innocent were correctly classified. However, the tests were administered in the same order to all subjects and this would be expected to influence the results.

There are a number of other studies using the guilty knowledge test and peak of tension tests to detect information (Gustafson and Orne, 1963, 1965; Miyake, 1978; Ben-Shakar, 1977; Ben-Shakhar and Lieblich, 1982; Kugelmass and Lieblich, 1968; Lieblich, Kugelmass, and Ben-Shakhar, 1970; Ohkawa, 1963; Timm, 1982; Stern et al. 1981; Kronbergerova and Dufek, 1969; Yamaoka and Suzuki, 1980; Dufek, 1969; Ohnishi, Matsuno, Arasuna, and Suzuki, 1976). These studies only use guilty subjects or subjects possessing information to be detected. Since there is no field analogue of this procedure, they will not be discussed. However, they are important in providing information on variables that influence detection of deception.

The guilty knowledge tests are consistent in demonstrating higher accuracy with innocent than guilty subjects. From 73% to 100% of the innocent subjects have been correctly identified, whereas from 53% to 95% of the guilty have been detected. It is not clear why there is such a range with guilty subjects. The number of stimulus repetitions (Balloun and Holmes, 1979), whether subjects are required to make a verbal response (Gustafson and Orne, 1965), and electrodermal lability (Waid and Orne, 1980; Giesen and Rollison, 1980) may be important. Whether the use of physiological measures in combination improves the accuracy should be investigated.

Guilty Information Tests - Reliability

Since guilty knowledge tests use an objective scoring system it would be expected to be highly reliable. The scoring of the peak of tension test is more subjective because there is judgment on the correspondence between maximum physiological activity and a critical item of information.

Chapter 2 Analysis of Research

Chapter 2

Analysis of Research

The previous chapter reviewed specific laboratory and field studies of the validity and reliability of lie detection. This chapter seeks to discuss some of the problems and issues of lie detection and to draw conclusions from the research literature.

Several observations can be made about the research. First, the research literature is burgeoning. There has been more scientific research conducted on lie detection in the last six years than in the previous 60 years. Second, the polygraph technique is a robust technique. Although there are numerous variables which affect its accuracy, it works better than chance in a wide range of testing situations including criminal investigations, intelligence operations, security screening, mock crimes in a laboratory environment, and even games in which a person denies what number he selected. Indeed, it is difficult to find experiments in which lie detection did not work better than chance. The technique obviously works well with many cultures because many of the studies were conducted in Poland, Israel, and Japan. Third, there is no such thing as a perfect study. Each study is flawed; sometimes in its design, often in its execution, and sometimes in the selection, analysis, and reporting of the data. Occasionally the shortcomings are so serious as to disqualify the study from serious attention. Fortunately, the various studies usually contribute to the mosaic of our understanding of the polygraph technique and the factors which affect its accuracy. Our understanding of these variables has expanded rapidly within the last five to ten years with the accelerating pace of research.

Nonetheless, it is difficult to estimate the precise accuracy of the polygraph technique in everyday applications because of the number of variables involved. There are a variety of polygraph applications, each with its peculiar problems

and issues: criminal investigation, intelligence operations, pre-employment screening, and aperiodic security screening of current employees. Within each of these applications are many subcategories. Polygraph accuracy is probably different for each of these settings; however, some general comments about error rates can be made.

There is some evidence to suggest that control question tests may be more accurate in detecting the deception of the guilty person than in verifying the truthfulness of the innocent person; that is, given equal numbers of guilty and innocent persons, more inconclusive results may occur with the innocent persons, and the number of false positives may outnumber the number of false negatives. Studies of this type have primarily found either more false positives than false negatives or essentially no difference in the false positive and false negative error rate. Since the variables determining whether or not a study showed differential error rates have not been determined, it is unknown whether a preponderance of false positive errors is characteristic of field polygraph usage.

Additional evidence that the guilty may be easier to detect was reported by Barland & Raskin (1975). They found that the absolute values of scores from guilty subjects in a mock crime were more extreme than the scores from the innocent subjects, indicating that the size of the reactions of the guilty subjects on the relevant questions were larger than those of the innocent subjects on the control questions. They also found that on numerically scored control question tests of criminal suspects, the scores of the suspects called deceptive by the examiner were more extreme than the scores of the suspects called truthful by the examiner (Barland & Raskin, 1976). Not all of those decisions could be verified, however.

It is of interest to note that in two of the studies reporting a preponderance

of false positives (Horvath, 1977; Kleinmuntz and Szucko, 1982), the errors occurred during the blind analysis of the polygraph charts. When verified cases were selected for study, no known errors were found regarding the polygraph tests as conducted by the original examiners.* This suggests the possibility that the clinical judgment of the examiner who is able to observe the subject's demeanor throughout the test may serve as a safeguard against false positive errors that may occur when the polygraph charts are interpreted in isolation from nonpolygraphic sources of data.

Peak of tension tests and especially the guilty knowledge test incorporate an extremely effective safeguard against false positive errors--the innocent person cannot determine which question is the critical question, and therefore cannot consistently react to it regardless of how nervous or fearful he is. Thus, with the peak of tension and guilty knowledge tests, virtually all errors are false negative errors. However, these tests can rarely be used in many applications of the polygraph. More research is needed to clarify the relationship between false positive and false negative errors with the control question test.

The base rates of truthfulness and deception within the specific population being examined can have significant effects on the proportion of false positive and false negative errors. The confidence to be placed in a given test outcome will vary according to the proportion of false positive and false negative results compared to the base rates of truthfulness and deception. As the base rate for deception decreases there will be an increase in the number of false negative errors and a decrease in the number of false positive errors. It should be recognized, even if there is doubt about a particular polygraph test outcome,

* Personal Communication, Frank Horvath, 1983; Personal Communication, Julien Szucko, 1983.

the test can nonetheless be useful. If, before the test, there is 1 chance in 1,000 that a person is deceptive, but afterwards there is a 50/50 chance the person is deceptive, the test has been helpful. This is particularly important to consider since the polygraph is supplemented by other investigative tools.

Another variable affecting the error rate is the type of issue to be resolved. The polygraph is believed to be most accurate when the subject denies having committed a specific physical act, such as a theft. Although there is, as yet, no research on this matter, several theories predict that the accuracy is reduced when the issue is vague or ambiguous, such as when the subject admits having shot the victim, but claims he had only intended to frighten, not kill, him. The polygraph issue is then one of intent, not the act. The extent to which the accuracy is reduced depends upon a number of factors unique to each case, such as the nature of the issue, the psychological makeup of the person being examined, the length of time elapsed since the event occurred, and the skill and experience of the examiner in formulating and defining the precise test questions.

The polygraph is believed to be most accurate when only one issue is to be resolved. When more than one issue must be included in the test, as often happens in many intelligence applications and in preemployment screening, the accuracy would be expected to be reduced. In the absence of research on this issue, it is not known whether the accuracy is reduced. The Correa & Adams (1981), Blum and Osterloh (1968), and the Barland (1981) studies suggest that there may be little if any decline in the accuracy of gross categorization of "entirely truthful" versus "deceptive to one or more questions", and that the decline in accuracy occurs in decisions regarding the precise question(s) to which the person was deceptive.

It is clear that the examiner's level of experience affects the accuracy of decisions. It may be that once a certain threshold has been reached additional experience results in only slight increases in accuracy, due to the inherent

robustness of the technique. However, the critical level of experience has not been determined. Within the government, the closely supervised internship and the quality control review process assists the intern in conducting his examinations properly.

The approach to decision making (clinical versus numerical analysis of polygraph charts) undoubtedly affects the accuracy of the decisions, but this factor has not been adequately researched. It may be that the clinical approach may be important in safeguarding against false positive errors while the numerical evaluation approach minimizes false negative errors. If this proves to be true, then a combination of both approaches should minimize both types of error. It would be a useful experiment to employ the clinical approach followed by a quality control review which incorporates numerical analysis of the charts to see if the combination is more effective than either approach alone. There is some evidence suggesting that when several polygraph examiners concur in the interpretation of the polygraph charts, the accuracy of the group decisions is higher than any individual opinion (Barland and Raskin, 1975). More research is needed to determine the most effective approach to decision making.

The number of polygraph charts obtained from a given subject may be related to the accuracy of the decisions. Generally, the more physiological information available to the examiner the more accurate his decisions are (Barland & Raskin, 1976). Beijk (1980) reported increasing accuracy rates on a card test with increasing numbers of polygraph charts. Studies reporting lower accuracy rates occur when the reviewers saw only one chart or only used a portion of the information available from the charts (Kleinmuntz & Szucko, 1982; Bradley & Janisse, 1981). The Kleinmuntz & Szucko study reported data from an experiment where the reviewers saw only one polygraph chart of a set, and had one of the lowest accuracy rates reported in any study. Multiple charts are taken in all federal applications of the polygraph.

One of the complicating factors in trying to use the available research to estimate the accuracy of the polygraph in real life situations is the fact that many laboratory studies are conducted by researchers who have not been trained in the field uses of the polygraph. Among the many differences that lack of field training generates is the relative lack of interaction between the researcher and the subject. There are many things that can cause responses to appear to indicate deception on the polygraph charts. These include extraneous noises, visual distractions, emotional complexes associated with some word in a question, and even random thoughts. The accuracy of the polygraph technique depends largely upon the ability of the examiner to control the testing environment so that extraneous sources of reactions are eliminated. "Lie detection" can properly be thought of as an inference based upon a process of elimination. One important way in which the field examiner determines whether a subject may be reacting to a question for some reason other than deception is to talk to the subject between charts to find out what was going through his mind on various questions. The questions can then be restructured based on the information the subject provides. This procedure would almost certainly reduce the number of innocent subjects classified as deceptive in laboratory tests. For example, Correa and Adams (1981) questioned the subject during the procedures and rephrased the questions based on the subject's replies. That study reported 100% accuracy in classifying the innocent and guilty subjects.

In addition to polygraph accuracy, polygraph utility also is important. Utility refers to the ability of a test or procedure to produce the results desired. If 100 tests are administered, it is desired that 100 correct decisions be obtained. Anything which reduces the number of desired results reduces the utility of the test.

As applied to the polygraph, utility has several aspects. In one sense it refers to the ability of the polygraph examiner to obtain useful information from a subject as a result of the test regardless of the actual accuracy of the polygraph

in terms of chart interpretation. During the interview conducted by the examiner immediately prior to attaching the subject to the polygraph, many subjects confess or reveal information they had not mentioned to previous investigators. The psychological "demand characteristics" of the polygraph situation are very powerful and should not be underestimated. Thus, even if the polygraph were not at all accurate, it would have a certain utility as long as people believed that it worked.

A second aspect of utility relates to the inconclusive rate. Even the most accurate test has diminishing utility as the inconclusive rate increases. Fingerprints, for example, have limited utility in investigations despite their extremely high accuracy because only occasionally can identifiable prints be recovered. In reference to the polygraph, it is necessary to distinguish between its utility and its accuracy. Utility in this respect refers to the total number of correct decisions out of the total number of cases; that is, the total number of cases minus both the inconclusive results and the errors. Accuracy, on the other hand, refers to the number of correct decisions out of the total number of decisions, after the inconclusive results have been set aside.

It is not possible to estimate the accuracy of the polygraph in DoD, at present, for the reasons discussed in this chapter. A great deal of government polygraph usage involves screening, but there is little evidence on the accuracy of the polygraph in screening situations. Two laboratory studies, Blum and Osterloh (1968) and Correa and Adams (1981) used procedures that are similar to the screening situation. Both correctly classified 100% of the innocent and 100% of the guilty subjects. Although there are not enough data to make definite statements about polygraph accuracy in screening, it can be said that there is no evidence to indicate that large numbers of people are misclassified.

There is more evidence on potential accuracy in DoD criminal applications. The Bersh (1969) field validation study was designed to estimate the accuracy of the polygraph technique in DoD criminal investigations.

This study found that in criminal investigations there was an agreement rate of 90% for guilty subjects and 94% for innocent subjects between the decisions of military examiners and the criterion of unanimous decisions of JAG attorneys who had reviewed the investigative dossiers minus the polygraph outcome. However, since the examiners knew some of the case information which was later forwarded to the panel of attorneys, these figures should probably be considered to be the upper limit for the accuracy of the polygraph technique in military criminal investigations in the early 1960's. Since that time the accuracy may have increased somewhat due to the higher selection standards, more thorough training (both the initial polygraph training and periodic seminars on advanced topics), and the establishment of quality control offices which review every polygraph case. For all of these reasons, it is likely that the quality and accuracy of federal polygraph examiners is higher than the polygraph profession generally.

Several laboratory studies have utilized numerically scored control question tests in mock crime situations with the physiological measures typical of the field. Thus, these studies use procedures which are similar to DoD procedures (Barland and Raskin, 1975; Rovner et al., 1978; Raskin and Hare, 1978; Podlesny and Raskin, 1978; Honts, 1982; Gatchel et al., 1983; Hammond, 1980). Bradley and Janisse (1981) used numerical scoring, but their scoring system differs from that used at DoD and they classified subjects as truthful or deceptive from only one physiological channel. Those experimental studies which used other test formats, such as the guilty knowledge test, or in which subjects were instructed to engage in countermeasures, were excluded from the data in order to make the results as applicable as possible to use the polygraph within DoD. Only those studies using numerical scoring are of interest because numerical scoring currently is used in DoD criminal investigations. The decisions of the original examiner were used for all studies except Rovner et al. (1978) because this study only reported results for a blind evaluation.

These studies correctly classify from 75% to 100% of the guilty subjects and from 57% to 100% of the innocent subjects. The mean correct classification rate weighed for number of subjects in the study is 90% for guilty subjects and 80% for innocent subjects.

There are numerous differences between mock crimes conducted in laboratories and real crimes. Some of these differences, such as a lower level of emotional involvement, probably made it harder to detect deception in the laboratory, whereas other differences, such as a greater homogeneity of the research subjects in terms of age, IQ, criminal history, etc., probably tend to increase the accuracy of laboratory results by permitting the polygraph procedure to be optimized for the subject population. Some of the more recent research has employed subjects recruited off the street or has used convicted criminals (e.g. Podlesny & Raskin, 1978; Raskin & Hare, 1978).

Since there is a limited amount of research data, it is important to consider the experience of professional investigators and quality control personnel in assessing field polygraph accuracy. The polygraph has been used as a lie detector in hundreds of thousands of criminal investigations from the moment of its inception in 1921. Police detectives and other professional investigators have worked closely with polygraph examiners, and have continued their investigations of countless cases after learning of the polygraph results. By their nature and training, investigators tend to be skeptical pragmatists. Yet, they rely upon the polygraph.

One of the functions of the various federal quality control offices is to follow up the cases, maintain records, and investigate causes of polygraph errors. Although it is impossible for all polygraph errors to be detected, the number of known errors in federally-administered polygraph examinations of criminal suspects is less than 1% of all polygraph examinations. One reason why the actual number of errors is unknown is that military prosecutors give a great deal of weight to

polygraph results when deciding whether to proceed with court martial proceedings. It is extremely rare for a person who has been found truthful by a federal polygraph examiner to be prosecuted. Consequently, false negative errors are not likely to be found. On the other hand, instances in which the subject is found deceptive are almost always investigated further and are usually prosecuted. Thus, the probability of false positive errors being detected by additional investigation is much greater. It is possible that the actual number of combined errors may be four or five times the number of known errors. This type of analysis would therefore suggest that the polygraph in criminal investigations is almost certainly more than 90% accurate, and that it probably is about 95% accurate.

Thus, the Bersh study, experience of investigators and quality control personnel, and mock crime laboratory studies give different estimates of the accuracy of control question tests in criminal investigations, ranging from about 80% to 95%. Until additional research is conducted we do not feel comfortable in specifying a precise figure for polygraph accuracy in DoD criminal investigations. Additional research also is needed to draw precise conclusions about polygraph accuracy for other DoD applications where different techniques are used, such as personnel screening and intelligence investigations. It is noted, however, that there are no data suggesting that the various polygraph techniques and applications at DoD have high false positive or high false negative error rates.

Chapter 3 Utilization of the Polygraph

THE POLYGRAPH IN THE FEDERAL GOVERNMENT

FEDERAL AGENCIES

The polygraph is currently used by the Department of Defense, United States Secret Service, Federal Bureau of Investigation, Postal Inspection Service, Alcohol, Tobacco and Firearms Administration, Drug Enforcement Administration, Central Intelligence Agency, United States Marshals, United States Customs Service, and the Department of Labor (Anti-Racketeering). These agencies often provide polygraph service to those federal law enforcement agencies and intelligence agencies and activities which do not have polygraph examiners. The Department of Defense elements which currently have polygraph programs are the United States Army Criminal Investigation Command, the United States Army Intelligence and Security Command, the Naval Investigative Service, the Air Force Office of Special Investigations, the U.S. Marine Corps Criminal Investigation Division, and the National Security Agency.

These fifteen agencies are represented on the Federal Interagency Polygraph Committee which meets quarterly, or more often if necessary. Although informal, the committee has achieved the adoption of minimum standards for the selection and training of federal examiners. The Committee has also established and operated an advanced polygraph seminar. Each year, the FBI hosts a one-week seminar attended by seventy or more federal examiners at its training facility at Quantico, Virginia. The program of instruction is planned and coordinated by a different federal agency each year, although in some years two agencies combine to put on the instruction.

Polygraph Examiners in the Department of Defense

All polygraph examiners in the Department of Defense must meet the minimum requirements set forth in DoD regulation 5210.48 plus those established in each service or agency. DoD examiners are college graduates, experienced investigators,

U. S. citizens, at least 25 years old, and specifically selected for their maturity and judgment. All have been subjected to a thorough screening examination with a polygraph, and have been the subject of an extensive background investigation. The basic polygraph training of military examiners is the twelve-week course conducted by the Army at the Military Police school, Fort McClellan, Alabama. All examiners then serve a closely supervised internship of six to twelve months.

Number of Examiners

As of September 1983, the Department of Defense had 153 certified examiners. That number does not include interns or those in basic training.

DoD CERTIFIED EXAMINERS

	<u>Army CID</u>	<u>Army INSCOM</u>	<u>Navy</u>	<u>Air Force</u>	<u>Marine Corps</u>	<u>NSA</u>	<u>Total</u>
1980	39	9	11	20	8	13	100
1981	42	9	12	20	6	30	119
1982	44	12	14	20	6	26	122
1983	50	15	15	40	6	27	153

Annual Polygraph Utilization

ANNUAL POLYGRAPH UTILIZATION (Number of Examinations)

	<u>Army CID</u>	<u>Army INSCOM</u>	<u>Navy NIS</u>	<u>Air Force OSI</u>	<u>Marine Corps CID</u>	<u>NSA</u>	<u>Total</u>
1980	4005	214	1317	1415	258	5676	12885
1981	3690	281	1185	1418	245	7418	14237
1982	3686	390	1337	2026	263	9672	17374

CRIMINAL INVESTIGATION

History

The Federal Bureau of Investigation and its predecessor, the Bureau of Investigation, made use of the polygraph technique occasionally from 1917 into the 1930's. In the mid-1930's Leonarde Keeler, who was then at the Northwestern University Crime Laboratory, trained an FBI agent in polygraph technique. A few years later he trained an agent from the United States Secret Service. The Army began to have agents trained as examiners in World War II, and when Leonarde Keeler established the first polygraph school in 1948 the federal government began to expand the use of the polygraph for criminal investigations. Today fifteen federal agencies or departments have polygraph programs.

Polygraph As An Aid To Investigation

In criminal cases, the polygraph is not used in the Federal Government as a substitute for investigation. Field investigation is usually conducted as far as circumstances permit before the polygraph plays a role. This requirement for prior investigation plus a prior and separate interview of all prospective polygraph subjects is mandatory in the Department of Defense, and may be waived by DoD officials only in the most unusual circumstances. Despite these restrictions, the polygraph is widely used in criminal investigations. Defense counsel in criminal cases frequently ask for exculpatory examinations.

The Army, which is the greatest user of the polygraph for law enforcement purposes in the Federal Government, has utilized the polygraph in 17% to 19% of its investigation of felony crimes during the last year and a half. Further, the Army has used the polygraph in 95% of its criminal investigations pertaining to crimes for which the maximum penalty is 15 years or greater. The effectiveness of this effort is illustrated by comparing the percentage of felony cases solved by the Army with the national rate. The Army solved 64.7% of their felonies while the

national average for the same period was only 19.5%. The use of the polygraph was one of the significant reasons for this extraordinary achievement.

The following tables depict the extent of use of polygraph examinations in criminal investigations in DoD, the pretest and post test confession rates, and the administration of exculpatory polygraph examinations which are requested by the suspect or his attorney:

CRIMINAL INVESTIGATION CASES

	<u>Army CID</u>	<u>Navy NIS</u>	<u>Air Force OSI</u>	<u>US Marine Corps</u>	<u>Totals</u>
1980	3990	1209	1336	255	6790
1981	3677	1049	1304	240	6270
1982	3665	1210	1745	261	6881

PRETEST ADMISSIONS OR CONFESSIONS

	<u>Army CID</u>	<u>Army INSCOM</u>	<u>Navy NIS</u>	<u>Air Force OSI</u>	<u>Marine Corps CID</u>
1980	12.5%	21.5%	11.5%	10.0%	1.9%
1981	10.3%	16.0%	8.9%	13.3%	3.7%
1982	10.7%	13.3%	11.0%	12.0%	3.4%

POST TEST ADMISSIONS OR CONFESSIONS OF DECEPTIVE SUBJECTS

	<u>Army CID</u>	<u>Army INSCOM</u>	<u>Navy NIS</u>	<u>Air Force OSI</u>	<u>Marine Corps CID</u>
1980	46.4%	19.4%	34.4%	35.6%	50.0%
1981	44.7%	27.7%	36.5%	33.3%	52.4%
1982	48.7%	51.3%	30.1%	31.2%	36.7%

EXCULPATORY EXAMINATIONS

(Requested by the suspect or his attorney)

Percentage of total cases in parenthesis

	Army	Navy	Air Force	Marine Corps
	<u>CID</u>	<u>NIS</u>	<u>OSI</u>	<u>CID</u>
1980	328 (8.2%)	239 (18.1%)	506 (35.6%)	39 (15.1%)
1981	349 (9.5%)	164 (13.8%)	466 (32.9%)	24 (9.8%)
1982	324 (8.8%)	209 (15.6%)	542 (26.8%)	31 (11.8%)

Criminal Investigation Case Examples

In 1980, an Army captain was killed when his parachute failed to open. Investigation revealed that the static line of the officer's parachute had been intentionally cut. As this crime occurred in an airborne division, the possibility existed that another parachute had been cut and that additional military members would be fatally injured during parachute jumps. As 160 parachute riggers had an opportunity to commit this crime they were all considered suspects. A decision was made to use the polygraph swiftly to determine the individual responsible for sabotaging the parachute. Five days after the death of the officer, use of the polygraph resulted in the identification of the person who sabotaged the parachute, who then confessed.

In 1982, a serviceman had been scheduled for trial for the rape and sodomy of a female military member. Based on a polygraph administered to the alleged assailant, who requested this test in exculpation, it was determined that he had not committed any sexual crimes with or upon the complainant. A subsequent polygraph examination of the female complainant resulted in a confession that she had made a false complaint because she was angry at some members of her former military unit. In this instance, the polygraph prevented an innocent man from going to trial for the serious crimes of forcible rape and forcible sodomy.

In June 1982, a female military person reported in a sworn statement that a male military person had assaulted her with intent to commit rape. A criminal investigation of the alleged incident failed to confirm or deny her complaint. Subsequently, the alleged incident became a security clearance matter and was reopened. The alleged perpetrator admitted "playing around" with the complainant, but denied any aspect of assault or attempted rape and he volunteered to undergo a polygraph examination. As a result of the examination, the male military person confessed to all the charges made by the complainant.

A civilian employee operating a government owned service (gas) station aboard a military installation was suspected of falsely reporting a \$1,800.00 robbery of the service station. HE was administered a polygraph examination which indicated deception. The employee then admitted stealing the \$1,800.00 himself and identified a local military member as an accomplice.

An individual was tried and sentenced to life in prison for the murder of a fellow soldier. This individual asked the Army to give him a polygraph examination for exculpation. The results revealed that he had not committed the murder. The investigation was reopened and the actual perpetrator of this crime was apprehended and the innocent soldier was released from prison. This matter would not have been resolved without the use of the polygraph.

The body of a nine-year-old girl was found in a trash dumpster on a military installation. An autopsy revealed the child had been sexually molested and the cause of death was asphyxia. Investigation revealed that an individual reported he had been robbed in the vicinity of the dumpster on the night in question. Further investigation revealed the complainant of the robbery had been in a bowling alley, which was the last place the girl was seen alive. The subject was offered a polygraph examination and he agreed. The examination results indicated deception on the issues of kidnapping and killing the girl. During the subsequent interrogation by the polygraph examiner, the subject made a full confession.

A U. S. military member was suspected of setting a fire aboard a ship that resulted in excess of one million dollars damage; he denied the charges. The military member was administered a polygraph examination which indicated deception, where upon he confessed to intentionally setting the fire.

An individual charged with Possession and Sale of Controlled Substance maintained his innocence during the investigation and requested an exculpatory polygraph examination. The examination was administered by an Army examiner and the individual was found to be truthful when he denied criminal participation in

the possession and sale of the controlled substance. The polygraph evidence was not considered during the pre-court martial hearings nor was it allowed to be introduced during the court martial proceedings. The individual was convicted and sentenced to serve time in prison. Following the conviction, the examiner continued the investigation and identified a civilian subject who confessed to the examiner that he committed the crime for which the military member was sent to prison. The examiner then coordinated efforts to obtain the release of the innocent military member from prison.

Criminal Conduct Revealed during Screening Examinations

A detailed analysis was made of the admissions made during the testing of 20,511 applicants for clearance and access at NSA from 1 October 1975 through 28 February 1979. There were 695 persons, 3.4% of the total, who admitted to the commission of a felony. In almost all of these cases the perpetrator had gone undetected. Among the few who had been caught, there were five or six who had been convicted, but who had falsified their forms to conceal the arrest, conviction, and time served. The admissions included murder, armed robbery, forcible rape, burglary, arson, embezzlement, hit and run driving with personal injury, thefts of expensive items or large amounts, smuggling and wholesale selling of illegal drugs.

In addition, 2,489 persons (12.1%) admitted to misdemeanors. These admissions included petty theft, simple assault, tax evasion, falsification of travel vouchers, malicious damage, indecent exposure, prostitution, selling illegal drugs and controlled substances for profit, deliberately writing bad checks, and other forms of fraud. Some of these cases may have been felonies. When there was doubt, they were statistically classified as misdemeanors. Not all of the misdemeanor admissions were disqualifying. The purchase, use or possession of illegal drugs was not listed in this category.

There were 8,383 (40.1%) who admitted to the use of an illegal drug on at least one occasion. There were 2,228 (10.9%) who admitted using amphetamines, barbiturates, hallucinogens and similar substances. There were 704 (3.4%) who admitted to the use of heroin, cocaine, or opium.

The statistical rates for the admissions of crimes and drugs have changed very little since that survey.

Examples of Crimes Admitted during Screening Examinations

The following admissions were made to NSA examiners during screening examinations. All are from FY 1982 cases:

An applicant for an engineering position at NSA, employed as an engineer by another government agency, admitted that his engineering degree was phony (he bought it through mail order from London for \$100). He also admitted that he shot and wounded his second wife (his prison term was not on his SPH) and his present wife is missing under unusual circumstances that he would not explain.

An applicant admitted to setting fire to the trailer that his ex-wife and child lived in. He had been questioned by the Michigan State Police as a suspect (attempted murder/arson) but denied it, and refused their offer of a polygraph test.

An applicant said he killed a young girl while in combat in Vietnam. He recognized her as a young girl when he shot her the first time, then for no reason he could give, he shot her several more times at close range. He called it murder. He also admitted stabbing a stranger in the face with a knife in an argument over some beer.

An applicant admitted to the forcible rape of his ten year old niece, which was never reported to the police.

An applicant for a position in which he would carry a firearm, admitted he had been charged with attempted murder but not tried for lack of evidence. He admitted firing his shotgun at six people, and hitting all of them.

An applicant admitted to firing a rifle into his estranged wife's home in an attempt to murder her. He fled Connecticut which had a current warrant outstanding for his arrest.

POLYGRAPH and the TRADITIONAL METHODS

The polygraph examination is part of the processing for clearance and access in NSA and CIA. Both are necessary for complete coverage. Although many agencies rely entirely on the background investigation, neither NSA nor CIA have considered authorizing SCI access without a polygraph in more than 30 years. During the Korean emergency in 1951, when resources were inadequate to conduct background investigations in a timely manner, NSA granted temporary access based on polygraph results and a National Agency Check, with full access on receipt of the background investigation.

1953 - Study of Polygraph/Background

In 1953, the National Security Agency conducted a study of the effectiveness of their emergency procedure of using the polygraph and National Agency Check.* That study considered 3,926 individuals who were processed for clearance and access by NSA (and its predecessor, AFSA) between 1 November 1951 through December 1952.

Of the 3,926 individuals concerned, 3,878 volunteered for a polygraph interview, and 48 refused to do so. These 3,926 individuals are further broken down into two categories: 2,902 who were interviewed prior to employment, and 1,024 who were recruited in the field and were interviewed after they were employed. Of those who refused to take a polygraph test, 41 were applicants and 7 were employees. There were 2,772 complete or partial background investigations reviewed during this study. The disparity between this figure and the total persons in the study is because, generally, when applicants revealed information during the polygraph examination which disqualified them from further consideration, no background investigation was initiated. Another contributing factor is that in the case of 64 employees and 209 applicants the reports of the

*NSA does not now authorize clearances on a polygraph and an NAC because PL 88-290 requires a full field investigation prior to clearance.

complete background investigations had not yet been received by the agency at the time of the study. Of the 1,017 employees who were interviewed a total of 455 of the tests indicated no deception and the employees gave no information requiring evaluation, while 569 indicated deception, gave information requiring evaluation, or they were physically or emotionally unfit for testing.

Of the 569 employees who required further security evaluation after the polygraph interview, 83 were terminated or permitted to resign in lieu of termination. In 65 of these cases the information which formed the basis for this personnel action was furnished solely as a result of the polygraph interview. Eight persons were terminated solely on the basis of information furnished from the background investigation, and 10 persons were terminated on a composite of the information from the two investigative methods. Of the 8 people who were terminated solely as a result of their background investigations, 3 made no admissions and appeared truthful on the polygraph test, while the other 5 made minor admissions or were unfit for testing. Altogether, 8.10% of all employees in the study were terminated, 78.3% on the basis of information from the polygraph interview, 9.6% on the basis of information from the background investigation, and 12.1% from a composite of the two. In addition to the 83 persons who were terminated, 248 persons covered by this study resigned voluntarily, for reasons unrelated to security, leaving 693, or 67.67% of this group remaining with the Agency.

The clearance of 76 employees was delayed pending receipt of background investigations. Six of these persons were later terminated on the basis of information provided by the background investigation, while in the other 70 cases no information was developed which was a bar to clearance. These persons had been held for the completion of a background investigation because they were unfit for testing, because they refused to volunteer for a polygraph test, or because the information resulting from the polygraph interview was not sufficient to enable

analysts to reach a security determination.

In the case of 2,902 applicants for employment, 41 persons refused to volunteer for a polygraph interview, leaving 2,861 who were examined with the polygraph. The tests of 1,251 applicants indicated no deception and the applicants gave no information requiring security evaluation. The remainder of the group of applicants, 1,651 persons (56.9%), required further security determination. Of these, 696 (24.3%) were disapproved for clearance on the basis of their admissions during polygraph examination. The 610 (21.3%) who gave minor admissions were evaluated as satisfactory for clearance. Of those 2,902 who applied for employment, 1,765 (61%) entered on duty. Some of the remaining 345 applicants were denied employment for reasons not related to security, others declined proffered employment.

There were 19 persons who were approved for hire on the basis of the satisfactory evaluation of their polygraph interview and a National Agency Check who were later terminated. Eleven of these persons had made no admissions and appeared truthful on their polygraph interview. In 3 cases the individual made minor admissions which were evaluated satisfactorily, and 2 cases involved those who had refused to volunteer for a polygraph test, but were nonetheless hired. A composite of the information from the background investigation and the polygraph test formed the basis for the termination of 2 persons, and the remaining individual was terminated on the basis of information developed during the polygraph interview of another applicant. The 17 persons on whom derogatory information was developed despite clear polygraph test results and no admissions or only minor admissions represent a false negative error rate of 0.6% in the testing of 2,902 applicants.

The use of the polygraph interview resulted in a considerable monetary saving in the security clearance of employees. In the group under study there were 696 applicants who were disapproved for hire and 65 employees who were terminated

immediately on the basis of information derived from the polygraph interview. Calculated on the basis of a nine-month clearance period at the GS-4 level, this group of individuals would have cost the government \$1,814,322 in salary alone while awaiting the completion of a background investigation. This figure is low, since the average salary of individuals awaiting clearances in the agency was above the GS-4 level and the average background time was more than nine months.

The study concluded that the background investigation is at its best in areas which are a matter of record, such as employment, education, naturalization, criminal history, etc., although even in these areas more pertinent information was developed through polygraph interviews. On the other hand, in personal matters not of record, the polygraph interview is far more reliable as a producer of information for security determination. In the polygraph interview this information is obtained directly from the subject. It does not, as with the background investigation, come from informants whose motivation and reliability are often unknown, and who, in a few cases where derogatory information is concerned, are vindictive.

The authors of the 1953 study observed that the primary justification for the use of the polygraph interview in personnel screening is not its financial savings to the government, but from the fact that a combination of a complete background investigation and polygraph interview provides more information for security determination than can be obtained from any other methods yet devised. Neither the polygraph interview nor the background investigation may be considered complete in itself. Each supplements the other; each improves the validity of security determinations.

1983 - Study of Polygraph/Background

In 1983, the National Security Agency conducted a pilot study comparing the information obtained from polygraph interviews where the examiner did not have any information from a background investigation, with background investigations where

none of the polygraph developed information reached the investigators. These "dual track" cases are unusual in that the investigation started well before the polygraph examination was conducted. Even so, some of the dual track cases had to be discarded from the study because information developed during the polygraph interview was totally disqualifying and the Defense Investigative Service was advised to stop the investigation before it was complete. Other cases were deleted from the study because information from the polygraph interview was given to DIS as leads for further investigation, because DIS had not yet closed their case.

The subjects whose files were reviewed in this study had already been evaluated as having acceptable information on their Personnel Security Questionnaires, and had received a polygraph interview sometime between 1 October 1982 and 31 January 1983. The polygraph interview and Special Background Investigation (SBI) were conducted independently, without any exchange of information. From a total of 248 randomly selected dual tracks cases, 194 cases were selected as meeting the criterion of complete independence, required by this study.

The results of the study revealed that 78% of all information used for evaluation was obtained from polygraph reports. The remaining 22% of information used for evaluation came from the Special Background Investigations.

In 113 cases (58%), the polygraph report was of particular value to the clearance adjudicator because it was the only investigative method to develop information for evaluation. By comparison, the SBI developed the only information in two cases (1%).

In addition, there were 22 cases (11%) in which the polygraph interview and the background investigation produced dissimilar but useful information, and 14 cases (7%) in which they developed similar information. There were 43 cases (22%) in which neither source produced useful information. Although polygraph interviews produced far more information than the background investigation, the contributions of the latter were vital, and often on matters not developed or even covered in

the polygraph interview. For example, credit is discussed in a full scope polygraph screening examination, but it is not the subject of a specific polygraph question because it is not practical to develop a dichotomous issue. Similarly, character deficiencies, alcoholism, a poor work record, and many other topics are often better learned from the observation of others, or from official records. In addition, the background investigation provides positive "whole person" information. The obvious conclusion was that both sources of information are productive, each from different sources, and both vital to the process of evaluation for clearance and access.

1974 - 1979 Polygraph Results from Cleared Military Personnel

A study was made of admissions made by applicants who were currently or recently in the military service who were cleared for SCI access and were in a cryptologic service. This table includes the 2,426 who applied in the fiscal years 1974 through 1979.

Admissions During Polygraph Interviews, FY 74 - 79from 2,426 Military Applicants

<u>Topics</u>	<u>Number of Persons making</u> <u>Admissions</u>	<u>Percentage</u>
Espionage	13	0.6%
Communist, Fascist or Terrorist Activity	25	1.0%
Divulgence of Classified Information:		
To family	135	5.6%
To others	132	5.4%
Secret Foreign Contacts	1	.04%
Soviet Bloc Travel	70	2.9%
Soviet Bloc Contacts	60	2.5%
Drugs		
Marijuana	1654	68.2%
Marijuana over 100 times	159	6.6%
Uppers, Downer, etc.	243	10.0%
Cocaine and Heroin	65	2.7%

Crimes

Misdemeanors	322	13.3%
--------------	-----	-------

Felonies	68	3.3%
----------	----	------

Alcohol Problems

Past	25	1.0%
------	----	------

Present	4	0.2%
---------	---	------

Deliberate Falsification

of the Security Form	152	6.3%
----------------------	-----	------

Serious Credit Problems	57	2.3%
-------------------------	----	------

Psychological Problems

Treatment	107	4.4%
-----------	-----	------

Suicide Attempted	14	0.5%
-------------------	----	------

In regard to the espionage admissions, most knew of or had strong reason to suspect others of engaging in espionage against the United States but had not reported it. A few were personally involved in espionage. See examples in this chapter.

1981 Polygraph Results from Cleared Military Personnel

A study was made of the 318 applicant cases for FY 81 in which the applicant was currently or recently in the military service in a cryptologic assignment with SCI access. In every case they had been the subject of a Special Background Investigation at some time in their career, and some had update investigations. All these investigative results had been evaluated as satisfactory for clearance and access. In some cases, the background investigation was in the security file available to the examiner. In the majority of cases it was not.

The table below lists only that information obtained during the polygraph examination which was not in the security file. The tables exclude all information obtained during background investigation, special investigations, and security interviews.

Admissions During Polygraph Interviews, FY 81from 318 Military Applicants

<u>Topic</u>	<u>Number of Persons making</u>	<u>Percentage</u>
	<u>Admissions</u>	
Espionage	0	0
Communist, Fascist or Terrorist Activity	1	0.3%
Divulgence of Classified Information:		
To family	13	4.1%
To others	10	3.1%
Secret Foreign Contacts	0	0
Soviet Bloc Travel	15	4.7%
Soviet Bloc Contacts	6	1.9%

Drugs

Marijuana	141	44.3%
Marijuana over 100 times	16	5.0%
Uppers, Downers, etc.	31	9.7%
Cocaine & Heroin	12	3.8%

Crimes

Misdemeanors	38	11.9%
Felonies	12	3.8%

Alcohol Problems

Past	7	2.2%
Present	1	0.3%

Deliberate Falsification

of Security Form	7	2.2%
Serious Credit Problems	14	4.4%
Psychological Problems		
Treatment	14	4.4%
Suicide Attempted	5	1.6%

Not all of the admissions listed above were disqualifying. The table does not disclose amounts or recency.

Note that the first major topics are similar to the topics usually included among the questions in counterintelligence polygraph examinations used for special access determinations.

1983 - Study of Interviews/Polygraph

One of the additional techniques sometimes used in processing people for classified access, particularly SCI, is a thorough security interview. For several years, NSA has conducted security interviews of incoming military personnel assigned to NSA. The interviewers are experienced Special Agents of NSA who are specifically trained for this work. The scope of their interview is very similar to that of the polygraph examinations given by NSA to civilian applicants for clearance and access.

A 1983 study at NSA reviewed the information obtained on 13 topics that are covered in both the security interview and the polygraph examination. A random sample of 100 security files containing military security interviews was taken from the list of those military persons who were interviewed between March 1981 and July 1983 on arrival at NSA. Another random sample of 100 cases for that period was taken from the security files in which there was a polygraph examination of a military assignee to NSA who took the examination as part of his processing for civilian employment at NSA.

In 49% of the cases, the interviewers obtained admissions from already cleared incoming military personnel which required security adjudication. In all, there were 87 specific admissions. The subject matter included ten of the thirteen topics covered in both the interview and the polygraph examination.

The polygraph examiners obtained admissions requiring security adjudication in 82% of the examinations of the already cleared military personnel who were seeking civilian employment. There were a total of 111 specific admissions. The admissions covered nine of the mutual topics.

A second analysis compared the security interview conducted on entry and the subsequent polygraph examination of those 100 military persons who applied for a civilian position at NSA. The security interview predated the polygraph examination by a range of a few months to three years. The polygraph examiner, of

course, had access to the interview report prior to the examination. What is interesting is that in these 100 randomly selected cases the examiner got information which predated the security interviews, on nine of the thirteen mutual topics. For example, in 31 cases (31%), the examiner obtained information on drug involvement that predated the security interview.

USE OF POLYGRAPH IN ARMY PERSONNEL SECURITY CASES

The U.S. Army has no general personnel screening program in which polygraph is utilized; however, the polygraph is utilized on a selective case-by-case basis. When the U.S. Army Central Clearance Facility (adjudications) receives a completed Background Investigation from the Defense Investigative Service in which there are areas of adverse information which remain unresolved, a polygraph examination may be requested to aid in the resolution. The polygraph examination is offered to the subject on a voluntary basis to aid in clarifying the areas not fully clarified by the investigation.

From February 1982 to September 1983, 180 such cases were referred for polygraph examination. Of the 180 cases, 67 subjects declined to take the examination. The following statistics summarize the results of those 113 cases in which the subject agreed to undergo a polygraph examination:

<u>Overall Results</u>			
<u>Nondeceptive</u>	<u>Deceptive</u>	<u>Inconclusive</u>	<u>No Opinion</u>
28(25%)	81(72%)	2(1.5%)	2(1.5%)

Where there was Deception Indicated (DI), relevant admissions were obtained in 71 cases (or 88%). The following is a breakdown by area of the DI admissions:

<u>Loyalty</u>	<u>Drugs</u>	<u>Sex Crimes</u>	<u>Other Criminal Acts</u>
16(20%)	40(49%)	11(14%)	14(17%)

Loyalty cases involve unreported contacts with foreign intelligence, unreported travel to denied areas, unreported contacts with persons within denied areas, association with communist/subversive organizations, and hostage situations.

Drug cases involve the use, possession, distribution, purchase or sale of illegal drugs.

Sex cases involved homosexuality, child molestation, and other aberrant sexual practices.

Other criminal cases involved allegations of criminal activity other than drug involvement or sex crimes, e.g. theft, embezzlement, fraud, etc.

Clearance Actions in Polygraph Cases

	<u>Deceptive</u>	<u>Nondeceptive</u>
Clearance Granted:	15 (24%)	24 (89%)
Clearance Denied:	36 (76%)	3 (11%)

Action was still pending in 19 DI cases and 2 nondeceptive (NDI) cases. Some of the DI cases were cleared up by minor confessions. The 3 NDI cases were denied access based on information developed during the polygraph examination.

The above figures indicate that adjudications were not based solely on polygraph results, but on the overall investigative file.

INTELLIGENCE, COUNTERINTELLIGENCE AND SECURITY

History

The instrumental detection of deception was first used in an investigation for the Army when a Secret Code Book was stolen from the Surgeon General's safe in 1917. The thief was detected among the 70 suspects tested, but he was not confronted. Followed from Washington, D.C. to New York City, he was apprehended in the act of passing the code book to a German agent. During World War I the Army trained a group of psychologists at Camp Greenleaf in lie detection techniques for counterintelligence purposes, but the Armistice was signed before they were put to use (Marston 1938).

Beginning in the 1940's, the polygraph was used at the Oak Ridge atomic facility for screening of employees. Operated entirely under contract by a private company, the program continued until 1951 when the security responsibilities then under contract were taken over by the government (Troville 1951). At the end of World War II at Camp Wetherill, New Jersey, the Army used the polygraph to screen 274 German prisoners of war who had been screened and selected for police leadership positions in the post-war German Government. A team of seven experienced examiners using RI Technique asked questions about Nazi party sympathy, communist sympathy, plans to engage in sabotage or subversion, activity for the Gestapo, SS, or SA, and serious crimes. The results cleared 156 Germans (57%) for senior police positions.

There were 8 inconclusive cases (3%) and 110 (40%) were not recommended. The examiners obtained admissions to Nazi Party membership by 24 candidates, and to SA by two and SS by one. Three were communists, of which one was a party organizer. Among the Nazis was the Party Treasurer from 1933 to 1938 (Linehan 1978).

In 1947 the newly formed Central Intelligence Agency began using the polygraph in support of operations and for specific investigations. They also began to use the polygraph in screening. By 1950, CIA was screening all applicants.

In May 1951, the Armed Forces Security Agency, predecessor of the National Security Agency, began using the polygraph to expedite the clearance processing of more than a thousand employees who had been hired but could not be cleared until the background investigation had been completed. Because of the Korean War, those investigations were taking from nine to eighteen months to complete. Accordingly, AFSA gave top secret access on an interim basis to employees based on a satisfactory polygraph examination and a National Agency check. Soon thereafter, AFSA-NSA began screening local applicants before they were hired and by 1953 was giving polygraph examinations to all applicants as a condition of employment.

During the years 1951 to 1983 a number of other special activities in the Department of Defense have used the polygraph for clearance and access determinations, and for operational purposes. Some of these polygraph operations were for specific projects of a limited duration, others were for long terms.

Acting on recommendations of a DoD Select Panel on Personnel Security, Frank C. Carlucci, Deputy Secretary of Defense, issued a Memorandum on 6 August 1982 which established an aperiodic counterintelligence polygraph program in the Department of Defense for those person who have SCI access. Mr. Carlucci also restored the Reinvestigation Program in DoD. The Carlucci polygraph program was implemented only in the National Security Agency. Other elements of DOD needed a revision in the DOD polygraph regulation before they could begin. The polygraph regulation was under revision when Congress expressed an interest in the proposed expansion of the polygraph in the Department of Defense. On March 11, 1983, the President issued National Security Decision Directive Number 84 "Safeguarding National Security Information," which was separate from the DoD action. Among the many security measures in NSDD 84 is the requirement to revise regulations and policies so that

employees may be required to take a polygraph examination in the course of investigations of unauthorized disclosures of classified information.

Congress will hold additional hearings, on NSDD84 and the polygraph, and has passed legislation which prohibits The Department of Defense from expanding the use of the polygraph beyond that existing on 5 August 1982 (the day before the Carlucci Memorandum), until 15 April 1984. NSA is exempt from this moratorium. The moratorium is to provide Congress with time for hearings and, if necessary, subsequent legislative action.

Specific Investigations of Espionage

In many cases, espionage activity is initially detected by other means, and the role of the polygraph examination is to find the guilty persons among those suspects developed by investigation. Not infrequently, the examiner will get admissions that are useful in furthering the investigation, or get confessions. Polygraph examinations were important to the successful prosecutions of Barnett, Kampiles, and Helmich. All received lengthy sentences for violation of the espionage statutes. And in each of these cases the polygraph examination was important to counterintelligence analysis and damage assessment.

Other espionage cases have been detected through screening operations conducted by the Department of Defense and the Central Intelligence Agency. Here, again, the examiners have obtained information in sufficient detail to initiate investigations, and in some cases have obtained confessions. It is important to note that in the screening situation, the applicant initiates the contact by applying for assignment, employment, or some form of access. The applicants know beforehand that they will be given a polygraph examination. The decision to take the examination and risk exposure must be a difficult one to make. There are several known cases in which agents have refused an assignment to seek employment where the classified access required a polygraph examination. This deterrent effect is illustrated by the refusal of Christopher Boyce to accept a job with the CIA as a courier between the United States and sites in Australia. He feared that the examination would reveal his espionage activities. He refused again for the same reason on a later occasion when his Soviet handler suggested he apply for work at CIA (Lindsey 1979). Likewise, Barnett opted not to apply for a staff position at CIA or NSA because he feared the polygraph test.

Prime Minister Thatcher (1982), in her statement on the report of the Security Commission about the Prime Case, said that "The Commission concludes that the polygraph is the only measure of which it could be said with any confidence that it would have protected GCHQ from Prime's treachery, either because it would have deterred him from applying to join or would have exposed him in the course of examination. In view of this and of the extreme gravity of the damage caused by Prime, the Government accepts the Commission's recommendation that a full and thorough pilot scheme should be carried out. The Commission recognizes that a polygraph examination would be seen by some as an unwarranted invasion of their privacy. But we are dealing with matters of the highest national security, and those who have access to the nation's most sensitive secrets must expect to be subject to the most rigorous vetting procedures."

Examples of Polygraph Cases - Espionage

A U.S. directed informant with access to the intelligence and security functions of a foreign power reported on an unidentified agent inside a highly sensitive U.S. communication center, and said that foreign government officials were reading U.S. classified communiques almost as soon as they were put on the wire. These communiques dealt with highly sensitive activities involving the Defense and State Departments, and the White House. Investigation led to five suspects. All were interviewed with sworn statements executed denying any activity on behalf of the foreign government. Investigation produced no additional indication of espionage activity. All five suspects agreed to undergo examination on the polygraph. Four of the five were determined by the examiner innocent of involvement. The remaining individual was reported as deceptive and subsequently confessed that he was, in fact, in the employ of the foreign power. He admitted furnishing far more information to the foreign power than was reported by the U.S.

informant. Much of the information he admitted providing to the foreign power was subsequently confirmed by the U.S. directed informant.

A U.S. military member was arrested after selling classified material to undercover federal agents following documented contact with the Soviets. The military member denied any additional contacts or any additional violations. He was administered a polygraph examination in which showed deception. He subsequently admitted to additional Soviet contacts and possession of additional classified information at his residence.

A U.S. controlled agent allowed himself to be recruited by a foreign intelligence service in 1978. From 1978 until 1982, the agent was met periodically by a foreign case officer, but never in the United States. In April 1982, the foreign intelligence service agreed to have a courier meet the agent in the United States. The courier was arrested when the meeting took place. He was tried and convicted. Throughout his interrogation following his arrest, he did not identify anyone else in the U.S. who might be engaged in espionage against the U.S. The convicted courier subsequently agreed to undergo a polygraph examination. The examination resulted in the courier furnishing the names of several persons in the U.S. known or suspected by him to be espionage agents for the foreign power he represented.

A military service received information from another agency that a military member was involved in espionage with a hostile intelligence agency. Subsequent investigation included a polygraph examination where the member admitted being involved in espionage against the United States for over a year.

A military service received information that an officer was involved in espionage. Subsequent investigation disclosed the suspect was a missile launch officer assigned to a western base. During one of his many trips he was arrested. Through his defense counsel, arrangements were made to grant immunity if the officer would disclose all compromises of classified information and pass a

polygraph examination to confirm the admissions. He was then thoroughly debriefed by counterintelligence personnel and said he had told everything. However, he made many additional significant admissions to the polygraph examiner before ultimately passing the polygraph test.

A foreign civilian who was formerly employed by U.S. Forces, was reportedly providing U.S. documents to a foreign intelligence service. That individual's daughter was married to a member of the U.S. military, and she was also implicated in the espionage activity. During an investigation of these allegations, three sources of information were examined by the polygraph and were determined to be truthful. Those examination results significantly aided in the development of the investigation. The host country police arrested the foreign civilian and his daughter for espionage. The American member of the military forces was interviewed and denied any knowledge of his wife or father-in-law being involved in espionage. He agreed to undergo a polygraph examination. The examination resulted in the member admitting that he knew his wife and her father were involved in espionage, and that he did not report that fact, even though he knew he was required to, because he wanted to protect his wife.

Espionage Detected During Screening Examinations

An applicant reacted to questions about involvement in intelligence for a Foreign government, espionage activity, intent to commit espionage against the United States, the accuracy of his security forms, and other related questions. After the tests, he admitted that his forms were not correct when he listed his academic affiliation during two long tours abroad, amounting to a total of eight years. He admitted that he was, during those years, a scientific advisor to the chief of a foreign military intelligence service. Although he admitted to the examiner that he might pass classified information to that service, he claimed that he had not done so. He had a Secret U.S. clearance before, between tours, and after his work abroad for that agency. He admitted that he was still in contact

with the foreign intelligence agency, and the subsequent test charts indicated consistent significant reactions to questions about giving or selling U.S. classified information to that service.

An Army Sergeant who had access to cryptologic information applied for a civilian position. During the polygraph examination, he reacted to various relevant questions. In the post-test interview, he admitted to various petty crimes and miscellaneous wrongdoing. The polygraph examiner noted continued specific reactions to relevant questions and when the Sergeant was reexamined several weeks later, the same situation continued. His access was withdrawn and an investigation opened. While that investigation was still in progress, he was found dead in his automobile. It was subsequently determined that he had been engaged in espionage on behalf of the Soviet Union.

A contractor employee seeking clearance and access reacted to polygraph questions concerning espionage. She told the examiner that her former husband was currently engaging in espionage against the United States for a foreign power. This had been going on for several years while her spouse held various positions in the government and with defense contractors requiring classified access. She gave very specific details on his espionage activities.

An applicant for employment reacted to questions about committing espionage, clandestine contacts, and related questions. He admitted that if employed, he would sell classified information to a foreign intelligence service if he could get enough money for it to live comfortably. The applicant had access to classified information while in the service. He continued to react to counterintelligence questions after his admissions.

An applicant described his various radical and Marxist connections including residence with a British Communist Party (CP) member while they both were students at a British university. The applicant visited the Soviet Union twice, both times with British student tours. He became particularly friendly with a young woman who

described herself as a student and part-time INTOURIST guide. She was his tour guide on the second trip. Applicant later wrote his tour guide telling her he had applied for employment with U.S. intelligence agencies. The Soviet tour guide came to the U.S., supposedly to visit relatives; contacted the applicant; then arranged to meet the applicant at an airport the day after his polygraph examination. He continued to react to counterintelligence questions during his polygraph testing.

An applicant said he had lived with a foreign national, who he said was an intelligence agent of a foreign power, while he was stationed in Berlin with U.S. forces. The applicant provided information about her activities including observation of a transmitter/receiver which was apparently for clandestine use. He did not report her activity to his superiors, despite the fact that she worked for U.S. military finance, because she paid for their apartment and he enjoyed living with her.

An applicant for employment who had lived abroad for several years overseas admitted that she was cultivated for several months by the host country intelligence agency and finally offered a position in which she was to travel to another country under a false identity, but as a U.S. citizen, then obtain employment in the target country (not the U.S.). She said that after two days of thoughtful deliberation, she declined. She had not previously reported this to U.S. authorities, and did so during the polygraph test only after reacting to counterintelligence questions.

An applicant who was about to retire from military service reacted to questions about intending to commit espionage against the United States. The applicant then described several visits to the Soviet Embassy to make arrangements to defect to the Soviet Union. The Soviet officials took copies of his documents, got extensive biographical information and when they learned of his pending applications for employment with U.S. intelligence agencies, encouraged him to stay in the U.S.

An applicant reacted to questions about knowing others engaging in espionage, intent on his part to engage in espionage, and to questions about the truthfulness of the information on his security forms. He admitted that he had been terminated from employment with a U.S. intelligence service after a year of training but before he was assigned to a position. After being fired he engaged in free-lance journalist work in the U.S. and abroad, with an Asian. During their work, the Asian told him he was an intelligence agent for his country. Although the applicant identified the agent by name, his location in the U.S., and his intelligence service, he would not give many details of their joint activities. After these admissions, he continued to react to questions about intending to commit espionage against the United States and the accuracy of his security forms.

An applicant admitted that a Russian, who stayed at his home for several days, asked for a tour of potential military targets in New Jersey. The Russian was allegedly with a Soviet delegation from the USSR Ministry of Health. The Russian took numerous pictures while the subject drove him to all the major bridges and petroleum plants in New Jersey, and to a nuclear power plant. The following year, he allowed two Soviet citizens from the Ministry of Health to stay in his home while they made tours during the daytime. They were vague about where they had been. The applicant provided additional information about other U.S. citizens who were assisting these Soviets in their non-medical activities.

One applicant, a frequent traveler to a Communist-bloc country, told the examiner of being drugged, and while drugged, an attempt was made to get him to engage in a homosexual act. Later, he was directly asked to work for the intelligence service of that country against the United States. He said he refused, but he did not report the entrapment attempt or the offer to any U.S. official.

Suitability Statistics

When considering for clearance and access those persons who have not yet had the opportunity to demonstrate their loyalty and trustworthiness, it is necessary to make judgments from their general conduct. Initial clearance is also the appropriate point at which undesirable people should be barred from access to classified information.

The extent and nature of admissions to criminal activity, including drugs, made by NSA applicants, has been set forth elsewhere in this report. Some other aspects of the admissions made during screening should also be reported .

Falsification of the Statement of Personal History or Personnel Security Questionnaire is reported only when it is deliberate, and the omission or false entry is significant. Among the 20,511 applicant polygraph examinations conducted at NSA from 1 October 1974 to 28 February 1979, there were 1,573 persons who falsified their security forms. That was 7.7% of the total number of persons examined. Omissions included criminal arrests and convictions, employment where their work was unsatisfactory or they were terminated, military service, relatives, subversive organizations, citizenship, education, foreign travel and numerous other matters. In many cases, but not all, an accurate record would have given information that would have been a bar to clearance and access, or to employment, or both.

Mental and nervous disorders were described by 1,306 applicants (6.4%). When appropriate, medical releases were obtained for use by those who conduct the background investigation. In addition, 297 (1.5%) admitted they had attempted suicide on at least one occasion. Because these applicants were also interviewed at NSA by a qualified clinical psychologist, an appropriate evaluation and prognosis was possible.

There were 2,820 persons (13.8%) who disclosed close contact with foreign nationals. Of particular interest were the 857 persons (4.2%) who admitted close

contact with Soviet bloc citizens, whose identities were not listed on the security forms. Many of these were contacts were made during foreign language training which developed into friendships, with contacts continuing through personal meetings and correspondence.

The examination includes questions about unauthorized divulgence. There were 2,153 persons in the 1974-1979 study who had prior access to SCI level material. Of those, 126 (5.9%) divulged classified information to members of their family and 129 (6.0%) disclosed classified information to others. The admissions, always checked by a classification authority, ranged from a one-time disclosure of confidential information to a member of the family to very serious compromises of extremely sensitive operations to outsiders, including foreign nationals. There were cases where the divulgence was not only serious, but frequent, and sometimes to total strangers. There were a few cases where persons who did not have official access to classified information heard it or saw it, and knowing it was classified, passed it on to other unauthorized persons. The most flagrant case of this kind was the son of a federal employee who was stationed abroad. The son had casual access to his father's office, and was able to get improper access to highly classified material in the office without others noticing. He then disclosed this information to the sons and daughters of foreign diplomats and other foreign nationals in order to impress them. The disclosures were serious, but he was never detected while doing this. Because some people do gain improper access to classified information, all applicants at NSA are asked about unauthorized disclosures.

SURVEYS OF EXAMINEES

1982 Air Force Survey

In 1982 the Air Force began a pilot polygraph project to conduct counterintelligence/security polygraph examinations of personnel who have access to extremely sensitive and highly classified USAF projects. After the project began, AFOSI began a survey of individuals who had voluntarily taken an examination in connection with that access. The questionnaire was answered in the reception room immediately following the examination out of the presence of the examiner. No identifying data was included with the replies. The 1,355 individuals responding to the survey gave their answers, totaled below in percentages, to each of the six questions. The "other" response includes answers such as "don't know" and no answers.

	<u>YES</u>	<u>NO</u>	<u>OTHER</u>
a. Were you offended, embarrassed, humiliated or degraded in any way during any part of the examination?	2.0%	98.0%	0
b. Do you think the examination was unfair in any way?	.9%	99.0%	.1%
c. Was there any objectionable or unwarranted invasion of your privacy during the examination?	.7%	99.1%	.2%
d. Do you feel counterintelligence security polygraph examinations enhance the security of your work environment?	94.9%	4.1%	1.0%
e. If you changed jobs and your new employer required a counterintelligence security polygraph examination, would you undergo the examination?	98.2%	1.5%	.3%
f. If an act of sabotage of espionage occurred and you were asked to take a polygraph examination to find the person who committed these acts, would you take a polygraph examination to resolved the matter?	99.2%	.7%	.1%

There are only two other surveys of employees who have taken polygraph examinations for security clearance and access, or have taken counterintelligence examinations.

1951 Survey of Atomic Energy Commission Employees

In 1951, Professor Paul V. Trovillo, a psychologist from Arizona State University, was employed to survey the use of the polygraph by a private contractor, Russell Chatham, Incorporated, at Oak Ridge, Tennessee. His survey covered the use of the polygraph on employees at the Atomic Energy Commission facilities who had sensitive access. All of the employees had a full field background investigation conducted by the FBI. The period covered by the survey was 17 February 1946 to 10 April 1951. (It is noted that none of the persons involved in the extensive Soviet atomic espionage network which operated during those years were among those who took polygraph examinations at Oak Ridge, Tennessee. Whether or not the polygraph deterred Soviet intelligence from an attempt to penetrate the Oak Ridge facility is unknown.) Trovillo reported that approximately 12,000 polygraph examinations were conducted in that period, including applicants for clearance by employees of contractors, and applicants seeking initial employment which required a Q clearance. In addition to the initial polygraph examination, each employee was subject to additional examinations. Depending on the sensitivity of access, some employees were reexamined every three or six months.

Trovillo personally interviewed a representative sample of employees who were in the retesting program. The employees included engineers, scientists, clerks, and chemical workers, including their union President. Of the 59 interviewed by Trovillo, only two (3.4%) expressed any resentment toward the polygraph tests (Trovillo 1947).

1956 Survey of NSA Applicants

A survey was taken of all applicants for employment at NSA who were processed

in January 1956. Of the 522 applicants who were surveyed, 361 were male, 362 were unmarried, 186 had a baccalaureate degree in a liberal arts field, and 206 had a degree in a scientific field. The survey covered these statements, with the responses indicated in percentages:

a. The polygraph examiner impressed me as being professionally competent and well trained.

1. strongly agree	58%		
2. agree	36%	agree	94%
3. undecided	4%	undecided	4%
4. disagree	1%	disagree	2%
5. strongly disagree	1%		

b. In my opinion, the polygraph as used by NSA is important to protect the security of the Agency.

1. strongly agree	38%		
2. agree	36%	agree	74%
3. undecided	19%	undecided	19%
4. disagree	5%	disagree	7%
5. strongly disagree	2%		

c. If for future reasons, I were asked to volunteer for another polygraph examination, I would:

1. be willing	91%
2. undecided	8%
3. refuse	1%

Surveys in Law Enforcement and Commerce

The Washoe County Sheriff's Department (Reno, Nevada) surveyed 85 applicants after their preemployment polygraph examinations. The results were tabulated (Putnam 1978) according to the responses to these questions:

1. Were you in any manner embarrassed, humiliated, or degraded by any part of the polygraph examination process? No: 92.9% (n=79), Yes: 7.1% (n=6).

2. In your opinion, was there any objectionable or unwarranted invasion of your privacy during the conduct of the polygraph examination? No: 100% (n=85).

3. Should you be hired, do you believe you will be more secure and comfortable in your work knowing the polygraph is used in personnel evaluation? Yes: 97.6% (n=83), No: 1.2% (n=1), no opinion: 1.2% (n=1).

Several very similar surveys of subjects of commercial screening have been made in recent years. Dr. B. A. Silverberg of Canada conducted three surveys of the subjects of polygraph examinations in 1980 asking the same questions used earlier in surveys by Ash (1973) and Buckley (1980). The questions were used again recently by Phannenstill (1983). A table shows the results of the six surveys. The answers which are other than the one listed for each item are mostly the opposite, but there are also a few unclear replies, or no answers at all to that item, in the remainder.

SURVEYS OF EXAMINEES OF COMMERCIAL TESTING

% (Number) Answered	Chicago & Denver		Buffalo &		Toronto		Milwaukee	
	Chicago Ash, 1973 n=241	Buckley, 1980 n=270	Western N.Y. Silverberg, 1980 n=102	Silverberg, 1980 n=217	Silverberg, 1980a n=115	Phannenstill, 1983 n=220		
1. Do you think the test was unfair in any way?	86% (208)	90% (244)	100% (102)	99% (216)	99% (114)	89% (195)		
2. Did the test or any part of it offend you?	91% (220)	87% (235)	97% (99)	98% (212)	98% (113)	87% (191)		
3. Do you think the test was an invasion of your privacy?	83% (200)	77% (209)	98% (100)	98% (213)	98% (113)	79% (173)		
<u>ANSWERED YES</u>								
4. If the occasion arose would you take a test like this as an applicant for a job?	96% (232)	94% (253)	93% (95)	95% (207)	93% (108)	94% (207)		
5. If a loss occurred at your company and you were asked to cooperate by taking the test to help find the person who caused the loss, would you?	97% (233)	97% (263)	96% (98)	96% (209)	97% (111)	96% (210)		

QUALITY CONTROL

Selection and Training

The quality control program within the Department of Defense begins with a detailed selection process of applicants for polygraph training. By regulation all applicants must be trained investigators with a minimum of two years Federal Government investigative experience and possess a baccalaureate degree. Applicants must also be a U.S. citizen, 25 years of age who have been the subject of a Special Background Investigation. The culminating application procedure has the qualified applicant undergo a suitability and familiarization polygraph examination prior to final selection. Polygraph training is always on a voluntary basis. Most important, agencies are careful to select only those agents who have demonstrated mature investigative and personal attitudes. Examiners must be able to conduct low key interviews as well as intense interrogations. They must be highly professional investigators who get along well with subordinates and superiors, as well as suspects, witnesses and victims.

The basic examiner training is accomplished at the Army Polygraph School, Fort McClellan, Alabama. This course is 12 weeks in length. The examiner learns various test techniques and receives instruction in psychology, physiology, pharmacology, semantics, and polygraph instrumentation procedures. Each trainee conducts at least of 45 to 50 practical exercise examinations prior to graduation. After graduation each examiner serves a six to twelve month internship under the direct supervision by a certified examiner of his respective agency.*

*NSA examiners take another two weeks of formal instruction in counterintelligence screening techniques before beginning their internship. NSA always includes students and experienced examiners from other DoD agencies in their course whenever requested; and on a space available basis, NSA trains examiners from other federal agencies.

In 1965 DoD initiated research to establish the validity of polygraph techniques and sought a means of monitoring field polygraph operations by having a qualified polygraph examiner conduct a "blind" analysis of the polygraph charts and allied documentation of another examiner. After a comprehensive study, it was concluded that blind analysis was a valid concept which would increase the quality of polygraph examinations. This concept led to the polygraph quality control programs now existing throughout the DoD. Under the quality control concept, supervisory polygraph quality control examiners, generally located at the headquarters level, review every polygraph examination conducted within their respective agencies. The review begins with a "blind" analysis of all polygrams collected by the field examiner. Once the "blind" analysis is performed, the reviewer then checks all related documents for correctness, completeness and technical accuracy. The specific areas reviewed are pretest interview, test construction, question formulation, chart patterns, chart markings, timing of questions, post-test interrogation, length of examination, and the polygraph report. Any notes, examiner comments and written admissions are analyzed and compared to the information contained in the formal report. The formal report is also checked for grammatical correctness and format prior to final approval.

At NSA, every case is completely tape recorded. In all cases where the subject of the interview has made admissions which will be reported and will require evaluation prior to granting clearance and access, the examiner reviews with the subject, in precise detail, the contents of his notes. While doing that, he makes an additional recording with a cassette recorder in full view of the subject, and has the subject audibly acknowledge the accuracy of his notes on each topic as it is covered. When the supervisor reads the examiner's report he must listen to this summary tape and certify on the report that the accuracy of the report has been verified by listening to the recording.

All Air Force counterintelligence/security examinations are audio and video recorded. Similarly, review of the audio and video tapes of these examinations by polygraph supervisors provides a detailed quality control check. In criminal cases, and specific issue cases, it is customary to take a witnessed, signed, sworn statement covering all admissions/confessions. A written statement containing a subject's admission/confession is documented in the final report as well as any refusal to undergo further questioning or polygraph testing.

Each agency keeps independent statistics and other personal information on the professional attributes, training qualifications and performance of each examiner. If an examiner's proficiency is found to be falling below accepted standards, supervisor determines if the problem is the result of an atypical series of cases, or a change in the examiner's methodology. Methodology changes are discussed with the examiner and, when appropriate, corrective action is taken.

Advanced training of at least one week or more is required once every two years for all DoD polygraph examiners. In addition, examiners generally receive several days of advanced training throughout each year by attending state and national polygraph association meetings and seminars. Advanced training is an integral part of the quality control program within the Federal Government as it assures the continuing professional development of each examiner.

US ARMY POLYGRAPH TRAINING FACILITY

The first use of the instrumental detection of deception was in World War I to recover a stolen code book. The U. S. Army began using the modern polygraph during World War II to screen German prisoners of war (POW) for post-war police assignments. Use of the polygraph spread quickly into criminal and counterintelligence investigations due to the success of the polygraph in resolving the POW cases. Some Army agents received individual instruction in polygraph techniques, and in 1948, the Army began sending special agents to the new Keeler Institute, in Chicago, Illinois, the only polygraph school in existence at the time. The Army's training needs soon exceeded the capacity of the Keeler Institute. Therefore, in 1951 the Army established its own polygraph school and attached it to the Military Police School at Fort Gordon, Georgia. Initially, the school was eight weeks in length and was used exclusively by the military services and the U.S. Secret Service.

In 1964, as a result of the Moss Committee Hearings about the use of polygraph in the federal government, the need for more extensive training was recognized and the school was lengthened to 12 weeks. In 1969, the school was lengthened to 14 weeks in order to accommodate training requirements by military and federal intelligence agencies in the personnel security screening arena. As use of the polygraph in the federal government increased, all federal investigative agencies, except the CIA, began training their examiner personnel at the Army school. Three agencies gave up their own training courses to take advantage of the excellent Army training.

In 1975, the Military Police School and the Polygraph School were moved to Fort McClellan, Alabama. During 1980, to accommodate the increased training requirements of the other federal agencies, the school was condensed to 12 weeks, permitting an increase of sessions from three to four classes per year.

Since the school's founding in 1951, the U.S. Army has trained over 2,000 military and federal civilian investigators to be polygraph examiners. Numerous training requests are received each year from state and municipal police agencies. While these requests cannot be accommodated in the basic training course, occasionally spaces are provided to state and municipal examiners in the advanced training course. The Defense Department plans to expand the Army Polygraph training facility to handle a maximum 24 students per class. Once expanded, the Army may be able to be more responsive to state and municipal police agency training requests.

There are two programs of instruction at the U.S. Army Polygraph School. The primary program of instruction is the Basic Polygraph Examiner Training Course, which is now twelve weeks in length and is conducted four times a year. Each class consists of 12 students and the student to instructor ratio is 2 to 1. This ratio provides the student with many hours of personal contact with qualified instructors during the learning process. Further, each student conducts over 40 polygraph examinations using a variety of testing techniques. This exposure to simulated testing and the instruction in multiple testing techniques has established the Army school as the leader in polygraph training.

The second program of instruction at the school is the Advanced Polygraph Examiner Course, which is three weeks in length. It provides experienced examiners with training in new instrumentation procedures and sophisticated testing techniques. These lessons are then incorporated in comprehensive practical exercises.

By regulation all DoD examiners must attend at least one advanced training course every other year. In practice, all examiners are involved in several short courses, seminars, and lectures every year, with a full course of a week or more at least every other year. Many get a week or more of formal training every year.

Advanced courses used by DoD examiners are quite varied. Most commonly used are:

U. S. Army Advanced Course	3 weeks
Federal Interagency Polygraph Seminar	1 week
National Security Agency Advanced Polygraph Screening Techniques Course	2 weeks
Federal Bureau of Investigation Advanced Polygraph Studies Program	4 weeks
Delta College Polygraph Workshop	1 week
University of North Carolina Polygraph Colloquium	1 week
University of Utah Polygraph Workshop	1 week
American Polygraph Association Annual Seminar	1 week

In addition, most DoD examiners are active in a regional or state polygraph association. These associations hold one or more weekend seminars each year, and a few conduct lengthy training courses. Several of the courses above and some of the state associations seminars offer college credits or continuing education units.

PRIMARY POLYGRAPH TESTING TECHNIQUES

The test formats taught at the Army Polygraph Training Facility include the following:

Zone of Comparison Technique (ZCT): The ZCT was developed by Cleve Backster in 1960. The ZCT is a modification of the control question concept developed in the 1940's and introduced two new safeguards against errors and inconclusive results. The safeguards are referred to as the sacrifice relevant and symptomatic questions. The sacrifice relevant question allows for dissipation of excessive general nervous tension or undue anxiety prior to the asking of the primary relevant questions. The symptomatic question provides a means to identify outside issues extraneous to the test issue(s) which may be of overwhelming significance to the examinee. Identification of such issues aids the examiner in structuring a reliable test that is not degraded by these external issues. The symptomatic question also helps the examiner to properly word the relevant questions. Backster's ZCT was adopted by the Army in 1961. It is the technique of preference for resolving criminal issues which are limited or easily defined. The ZCT, as used by the Army, permits the asking of questions pertaining to two primary issues in the body of the test. It also has a "SKY" phase utilized on one or two charts which helps determine the degree of the examinee's criminal knowledge, if any, of the primary relevant issue. Because of its demonstrated accuracy (Berish 1969), it has become one of the most popular and widely used techniques in government law enforcement.

Modified General Question Technique (MGQT): The MGQT was adapted by the Army in 1956 from the 1953 Reid control question test. It is the technique of preference in criminal investigation testing where there are multiple issues within a singular case to be resolved, as it allows the examiner to address primary

involvement, secondary involvement, guilty knowledge and evidence connecting issues. The MGQT is one of the more popular examination formats in government law enforcement because of its flexibility. In addition to being able to address multiple issues, examiners can shift the relevant questions on one or more charts by collecting a mixed series questioning sequence. The Army's MGQT utilizes control questions which are separated from the relevant issues by time and/or place.

Relevant/Irrelevant (R/I): The R/I techniques were developed in the 1920's and 1930's, and the Keeler version became the first standardized polygraph question technique. R/I techniques were used exclusively within the U. S. Government until 1956, when a control question technique was introduced. R/I techniques are used extensively in intelligence screening examinations and occasionally in criminal investigations. R/I techniques in the government employ the use of irrelevant questions interspersed among relevant questions, plus additional special purpose questions, including one or more control questions. The R/I technique is the technique of preference for people who decline to answer broad crime-related questions necessary for the development of control questions as they are used in ZCT and MGQT, as well as these persons who are familiar with criminal polygraph testing procedures.

Peak of Tension (POT): The POT was standardized by Leonarde Keeler in the 1930's and is used in conjunction with virtually every type of testing technique known to polygraph. Essentially, the POT is a guilty knowledge test where the examinee's knowledge of a particular item within the overall testing issue can be detected. This test is frequently used to find weapons, bodies, property, etc. It can be a known POT, wherein the examiner and the guilty subject know the key item and the innocent person does not, or a searching POT, where the key item is only known by the guilty subject and not by the examiner.

Counterintelligence Screening Test (CIST): The CIST is a general question test used to cover multiple issues pertaining to intelligence or counterintelligence

activity. If an examinee shows response to relevant questions during the course of an examination, the issues to which the examinee showed responses will be completely explored through the use of other polygraph techniques or through interrogation. Each relevant question will be asked at least three times.

Other Techniques: The techniques described above are among the techniques utilized in the federal government. There are a number of other techniques used in the law enforcement and commercial polygraph field. Among the more popular are those developed by John E. Reid, Richard O. Arther, and Lynn Marcy. Although some civilian schools teach these techniques, there are other civilian schools which teach the Army techniques, or teach both Army and other techniques. Government examiners are given familiarization lectures on the more common other techniques so that they will understand them; but they are not expected to use them.

Chapter 4 Summaries of Research

Field Abstracts

Gordon H. Barland and David C. Raskin. "Validity and Reliability of Polygraph Examinations of Criminal Suspects." Report No. 76-1, Contract 75-NI-99-0001. National Institute of Law Enforcement and Criminal Justice; Department of Psychology, University of Utah, Salt Lake City, Utah, March 30, 1976.

Procedure

One hundred-two criminal suspects (92 independent cases) were subjects in a study to determine the validity and reliability of the polygraph in a field setting. Most of the subjects were under arrest at the time of the examinations. No victims, witnesses, or persons involved in civil suits were included.

Eighty-four males and 18 females (age range, 15 to 54 years; education range 3 to 16 years) received polygraph tests where abdominal or thoracic respiration, skin resistance, and cardiovascular activity were recorded. Location of test and type of instrument varied with the subjects. It is noted that 34 subjects were tested with a polygraph in poor condition. Experience of the polygraph examiner is not discussed.

During a pretest interview the examiner obtained a background history, administered several scales from the MMPI, determined what the subject knew about the crime, and formulated the test questions with the subject. The federal modification of Backster zone-comparison control question test was used in 101 cases. One case used the U.S. Army 1966 version of the Reid technique. A minimum of three charts (10-13 questions per chart) were run; if no decision was made, additional charts were run up to a maximum of six charts. Charts were numerically evaluated at time of test by scoring each of three pairs of relevant and control questions from +3 to -3 based on magnitude of the differences. Truthful was indicated by a total test score of +6 or greater, deception by -6 or lower, and inconclusive between +5 and -5. In some cases the examiner's decision was based on more information than the numerical evaluation of record.

Each case was independently reviewed by a panel with no knowledge of the polygraph result. Evidence on each case was collected by untrained investigators and summarized by technicians with no knowledge of the case. All available documentation was included except polygraph and judicial outcome. If a subject confessed as a result of the polygraph test, the confession was reported but the source of the confession was not. If the subject had pleaded guilty, the circumstances of the guilty plea were included. The panel consisted of two prosecuting attorneys, one judge, and two defense attorneys. Panelists were to disregard legal technicalities and rules of evidence, and judge guilt and innocence based on the facts. Each member independently reviewed each case and made one of five possible decisions; (1) definitely innocent; (2) probably innocent; (3) undecided; (4) probably guilty; (5) definitely guilty. These decisions were converted to numbers on a 5-point scale by the examiners. Scores were summed for each case and total panel score was compared to polygraph outcome.

The following table demonstrates the agreement between panel decision and examiner decision when +1 or greater panel score reflected innocence, -1 or lower panel score reflected guilt, and a 0 panel score reflected inconclusive. The ten examinations that were not independent were eliminated in this analysis.

<u>Examiner Decision</u>	<u>Panel Decision</u>			
	<u>Deceptive</u>	<u>Guilty</u>	<u>Innocent</u>	<u>Inconclusive</u>
		47	14	5
	<u>Truthful</u>	3	6	4
	<u>Inconclusive</u>	6	5	2

Omitting inconclusives there was agreement in 76% of the cases. Increasing the panel score required for agreement increased rate of agreement (also note there would be an increase in inconclusives) and it fluctuated around 87%.

The polygraph examiner's judgment also was compared with various other types of criteria. The following results were obtained when panel decisions were based on a majority of the panelists making a decision in the same direction.

<u>Examiner Decision</u>	<u>Panel Decision</u>			
	<u>Deceptive</u>	<u>Guilty</u>	<u>Innocent</u>	<u>Inconclusive</u>
		43	9	14
	<u>Truthful</u>	0	5	8
	<u>Inconclusive</u>	4	3	6

When panel results were compared to blind numerical evaluation of charts similar results were obtained:

Polygraph examiner's decisions also were compared to judicial outcome in 41 cases where the jury did not know a polygraph test had taken place and are as follows:

<u>Examiner Decision</u>	<u>Judicial Outcome</u>		
	<u>Guilty</u>	<u>Innocent</u>	
	<u>Deceptive</u>	30	6
	<u>Truthful</u>	0	1
	<u>Inconclusive</u>	3	1

Results were similar with blind numerical chart analysis.

The relationship between judicial outcome and panel majority decision is in the following table:

Akiva Ben-Ishai. "Some Remarks on Polygraph Research." Paper presented at the Ninth Annual Meeting of the American Embassy of Polygraph Examiners, Chicago, Illinois, August 1962.

Procedure 1

A psychology student, assigned to the Israeli Police Department, determined the relationship between polygraph examiner decision and case outcome for those cases whose results had been determined by other means. It should be noted that the description of the procedure is incomplete.

Results

The psychology student's evaluation of guilt or innocence from the records was consistent with the polygraph results in 94% of the cases. It is not possible to determine from the presented data the percentage of guilty subjects and the percentage of innocent subjects correctly classified.

Procedure 2

The student was given all the charts from ten criminal cases in which the guilt or innocence had been definitely established by independent means. The student measured the amplitude of skin resistance responses to each of the control questions and to each of the relevant questions. He then averaged all the responses to the control questions and averaged all the responses to the relevant questions. If the average amplitude of the control question responses was greater than the average amplitude of the relevant question responses, the subject was said to be nondeceptive. If the average amplitude of the responses to the relevant questions was greater than that of the control questions, the subject was said to be deceptive. The chart analysis did not include respiration or cardiovascular recordings, and was limited to this unusual method of averaging electrodermal responses for analysis.

Results

All ten subjects were correctly classified as guilty or innocent.

Philip J. Bersh. "A Validation Study of Polygraph Examiner Judgments."
Journal of Applied Psychology 53 (1969): 399-403.

This was a report of research for the Department of Defense Joint Services Group on Lie Detection Research, United States Army Behavioral Science Research Laboratory, Washington, D.C., April 1968.

Procedure

The lie detection judgments of polygraph examiners in criminal investigations conducted by the military services were validated against unanimous guilt-innocence decisions by a panel of four Judge Advocate General (JAG) attorneys. The panel of lawyers had access to the complete investigation file, except that all references to the polygraph examinations were removed. Cases were selected at random from those conducted from 1963 to 1966 with some restrictions. Cases involving confession were ruled out because some polygraph examiners could have made their judgment of deception after the subject had confessed. No cases where the examiner reported inconclusive results were included. Cases were selected so that there were about an equal number of General Question tests and Zone of Comparison tests, and with each type, an equal number of judgments of deception indicated and no deception indicated. Examiners did have access to a subject's case file, but the authors maintain the file was not complete and that polygraph tests are only given when there is real doubt about a person's guilt or innocence. The polygraph examiners' judgments were subjective decisions regarding subjects' guilt or innocence.

A total of 326 case files were submitted to the panel. Each member of the panel was initially required to judge whether a file contained sufficient evidence to warrant a decision of guilt or innocence. Files with inadequate evidence were eliminated from further consideration, and 157 remained in which all four decisions were unanimous. The attorneys were given explicit instructions to disregard all legal technicalities and to judge each case solely on the evidence contained in the file.

Results

Results are presented in the following table (cases are those involving unanimous panel judgments):

		<u>Panel Judgment</u>			
		<u>General Question Test</u>		<u>Zone of Comparison Test</u>	
		<u>Guilty</u>	<u>Innocent</u>	<u>Guilty</u>	<u>Innocent</u>
<u>Examiner Judgment</u>	<u>Guilty</u>	31	4	34	3
	<u>Innocent</u>	1	32	4	48

The examiner and the panel agreed in 92% of the cases. When polygraph examiner judgment was compared with panel judgment for cases where a majority of the panel agreed on guilt or innocence (instead of unanimous decisions), there was agreement by the examiner and the panel in 75% of the cases.

M. E. Bitterman and F. L. Marcuse. "Cardiovascular Responses of Innocent Persons to Criminal Interrogation." American Journal of Psychology 60 (1947) 407-412.

Procedure

A theft of \$100.00 took place in one of the rooms of a campus dormitory at Cornell University. It was recognized that the theft could have been committed either by one of the men in the dormitory or an outsider. Two professors of psychology who had no polygraph training but did have some books and articles on the topic, conducted polygraph examinations of the 81 men who lived in the dormitory, as an aid to the police investigation. Respiration and cardiovascular activity were recorded. They were unable to devise a method to evaluate the pneumograph pattern. The first stage of testing used a fixed sequence relevant-irrelevant technique of seven questions. It opened with two irrelevant questions, then a relevant question, an irrelevant question, and three relevant questions. The last relevant question was a general one, asking the subject if he had answered all the questions truthfully. The whole pretest interview and one chart took ten minutes.

Results

Subjects were placed in 1 of 4 categories based on their responses to relevant and irrelevant questions (reliability data on this classification scheme is presented): 1) Insignificant response to all items (N=31); 2) Moderate reactions to all items (n=23); 3) Extensive reactions to all items (n=20); and 4) Larger reactions to relevant than irrelevant items (n=7). Subjects in groups 3 and 4 were retested from 1 to 5 times using relevant-irrelevant tests; with retest their patterns became similar to those of groups 1 and 2. Supplementary peak of tension tests indicated nothing significant. Thus, the authors concluded that none of the men had committed the theft. Their conclusion was supported by other evidence (not specified) which indicated that a person who was unavailable for examination was guilty. Thus, this study classified 100% of the innocent subjects correctly.

Avital Ginton, Netzer Daie, Eitan Elad, and Gershon Ben-Shakhar. "A Method for Evaluating the Use of the Polygraph in a Real-Life Situation." Journal of Applied Psychology (1982) 67(2): 131-137.

Procedure

Twenty-one males (mean age 29.2 years) from the Israeli police participated in a study involving detection of deception in a real-life situation. Subjects were administered pencil and paper tests presented as part of a police course. The answer sheet for one of the tests had a hidden chemical page that recorded what the subject had written. The chemical page was removed, the answer sheet returned, and subjects were asked to score their own tests. Seven subjects actually changed their answers as they were scoring the test. This was determined by comparing the scored answer sheets with the answers indicated on the chemical page.

After several days subjects were told they were suspected of cheating and were offered an opportunity to take a polygraph test. It was made clear that their future could depend on test outcome. All 21 subjects initially agreed to the test; however, 1 guilty subject did not arrive for the test, 3 guilty subjects confessed, and 1 guilty and 1 innocent subject refused to take the test. Thus, 13 innocent subjects and 2 guilty subjects took the polygraph examination.

Testing was done at police laboratories by 7 interrogators with at least one and one-half years experience as professional polygraphers. Respiration, electrodermal activity, and cardiovascular activity were recorded. Each interrogation involved 3 examiners - 1 interrogator, 1 who later blindly evaluated the charts, and 1 who observed from another room. A control-question test with 9 questions, 2 relevant questions, and 2 control questions was administered. There were 3 repetitions of the 9 questions and a card test after the first series. The polygraphers did not know subject status or the proportion of guilty subjects. Two months later charts were blindly re-evaluated by 8 examiners (5 of the original examiners and 3 additional examiners). Decisions of deception and no deception were based on both global subjective judgments and a numerical field scoring technique. The numerical procedure involved scoring each relevant and control pair from +3 to -3 and summing across responses and charts. Scores between +5 and -5 were considered inconclusive. Global subjective judgments were based on different data depending on the examiner's role in each case. The blind evaluators used the polygraph charts, the observer used subject behavior, and the interrogator used both the charts and behavior.

Results

The following table presents results of subjective judgments of the 3 examiners initially involved in the polygraph testing of the 15 subjects:

		<u>Guilty</u>	<u>Actual</u>	<u>Innocent</u>
	<u>Interrogator</u>			
	<u>Guilty</u>	2		2
	<u>Innocent</u>	0		11
	<u>Inconclusive</u>	0		0
<u>Predicted</u>	<u>Blind Evaluator</u>			
	<u>Guilty</u>	1		3
	<u>Innocent</u>	1		7
	<u>Inconclusive</u>	0		3
	<u>Observer (behavior only)</u>			
	<u>Guilty</u>	0		2
	<u>Innocent</u>	2		11

The original interrogator who based his decision on polygraph charts and behavior was most accurate, correctly classifying 100% of the guilty subjects and 84.6% of innocent subjects.

Results for numerical scoring are presented in the following table:

		<u>Guilty</u>	<u>Actual</u>	<u>Innocent</u>
	<u>Interrogator</u>			
	<u>Guilty</u>	1		1
	<u>Innocent</u>	0		6
	<u>Inconclusive</u>	1		6
<u>Predicted</u>	<u>Blind Evaluator</u>			
	<u>Guilty</u>	1		1
	<u>Innocent</u>	0		5
	<u>Inconclusive</u>	1		7

Numerical scoring increased the number of inconclusives, but did not change the accuracy of the interrogator if inconclusives are omitted. The blind evaluator did have higher accuracy with numerical scoring (omitting inconclusives). The authors note that the blind evaluator was under considerable time pressure for the subjective scoring and suggest that this factor may have influenced accuracy.

The 8 polygraphers who reanalyzed the data did so using both subjective and numerical scoring. There were 16 decisions for each type of scoring for guilty subjects (2 guilty subjects, 8 examiners) and 104 decisions for each type of scoring for innocent subjects (13 innocent subjects, 8 examiners). Results are in the following table:

		<u>Guilty</u>	<u>Actual</u>	<u>Innocent</u>
	<u>Subjective</u>			
	<u>Guilty</u>	15		18
	<u>Innocent</u>	1		84
	<u>Inconclusive</u>	0		2
<u>Predicted</u>	<u>Numerical</u>			
	<u>Guilty</u>	11		8
	<u>Innocent</u>	0		38
	<u>Inconclusive</u>	5		58

Eugene C. Edel and Jacob Jacoby. "Examiner Reliability in Polygraph Chart Analysis: Identification of Physiological Responses." Journal of Applied Psychology (1975) 60(5): 632-634.

Procedure

Forty randomly-selected polygraph cases from preemployment screening examinations conducted by a federal agency were used to determine reliability of polygraph chart analysis. There were 10 polygraph examiners ranging in actual interview experience from three months to ten years. Each examiner was the examiner in 4 of the cases and was a blind rater (with no knowledge of interview information) in 8 cases. Thus, each case was analyzed by 1 examiner and 2 blind raters.

The forty cases contained 2,530 relevant, irrelevant, control, and overall-truth questions. The relevant-irrelevant technique was used and several topics were covered in each examination. Examiners were required to state whether or not there was a significant physiological response in the cardiosphygmograph, electrodermal, and respiration channels to each questions, requiring 7,590 decisions. The federal agency which conducted these examinations does not report deception or no deception in screening cases. It only reports either that there are consistent, significant physiological reactions to specific relevant questions, or there are no reactions to relevant questions. Agreements between examiners and raters were based on the responses following each question in each of three recording channels. Thus, if both judges said there were significant physiological reactions to the question in the electrodermal channel, and no reactions in cardiovascular and respiratory channels, this would constitute three agreements.

Results

The percentage agreement between examiner and both raters was 96% for cardiovascular, 95% for electrodermal, and 96% for respiration, with the average agreement across physiological channels 96%. Percentage agreement between raters was 96% for cardiovascular, 91% for electrodermal, and 96% for respiration with an average agreement across channels 94%.

Eitan Elaad and Esther Schahar. "Polygraph Field Validity," in Israel Nachshon (ed.) "Scientific Interrogation in Criminal Investigation." Selected papers presented at the First National Conference on Scientific Interrogation in Criminal Investigation held at Bar-Ilan University, Ramat-Gan, Israel, 3-4 November 1976. Reprinted in a Special Issue of Crime and Social Deviance 6(1) (Spring-Summer 1978): 4-5.

Procedure

Of all the cases conducted by the polygraph examiners in the Scientific Interrogation Unit of the Israel Police during the years 1973 and 1974, 184 were confirmed by conviction or confession. The classification of these cases as deceptive or truthful was determined by the utilization of the polygraph.

Results

Of the 184 cases that were confirmed, there were 10 polygraph results that were inconclusive (18.4 percent). There were 6 errors (3.3 percent), 5 false negatives and 1 false positive. Eliminating the inconclusive reports, and assuming the conviction or confession was correct when there was disagreement, polygraph tests correctly classified 96.6% of the guilty subjects. The authors note that polygraph results could have influenced the confessions and convictions, the sample was not random, and false confessions and convictions could occur.

Robert H. Edwards. "A Survey: Reliability of Polygraph Examinations Conducted by Virginia Polygraph Examiners." Bureau of Forensic Science, Department of General Services, Commonwealth of Virginia, July 31, 1981, Richmond, Virginia, 62 pp. Reprinted in Polygraph 10(4) (December 1981): 229-272.

Procedure

The study, conducted for the Commonwealth of Virginia, involved a survey of fully licensed Virginia polygraph examiners. Twenty-four examiners of the Virginia State Police and 38 examiners with local law enforcement agencies responded (77.5% of the total law enforcement sample). The response rate for private firms was so low it will not be reported.

The survey questionnaire had six questions including questions about the total number of cases, the conclusion of each case (truthful, deceptive, inconclusive), the number of cases verified as truthful and verified as deceptive, and the number of those cases in which later information indicated the examiner was in error (false negatives, false positives). Only specific issue examinations were included.

Results

The 62 law enforcement examiners included in the survey conducted 2,433 specific issue examinations in 1980. Of these, 1,113 (45.7 percent) were reported as truthful, 983 (40.4 percent) were reported as deceptive, and 337 (13.9 percent) were reported as inconclusive. Excluding the inconclusives, 53.1 percent were reported truthful and 46.9 percent were reported deceptive.

The following table illustrates the polygraph examiners' determinations and the dispositions of the 959 cases that were verified by means such as conviction, acquittal, confessions, etc.

<u>Examiner's</u> <u>Decision</u>	<u>Guilty</u> <u>Innocent</u>	<u>Case Disposition</u>	
		<u>Guilty</u>	<u>Innocent</u>
	<u>Guilty</u>	587	9
	<u>Innocent</u>	7	356

Thus, 98% of the cases reported by the examiner as truthful were verified as truthful and 98.5% of the cases reported by the examiner as deceptive were verified as deceptive. It should be noted that case disposition was not necessarily independent of the polygraph examiner's decision.

Frank Horvath. "The Effect of Selected Variables on Interpretation of Polygraph Records." Journal of Applied Psychology 62 (2) (1977):127-136.

Procedure

Ten field-trained polygraph examiners from law enforcement agencies (5 having more than 3 years experience and 5 having less than 3 years experience including some interns) judged a sample of polygraph records from 112 criminal suspects from independent cases investigated from 1969-1972. The records were sampled in stages so that there were an equal number of verified and unverified cases, truthful and deceptive subjects, and crimes against people and crimes against property. Thus, there were 8 categories and 14 sets of records in each category. Verified cases were those where a confession had been obtained and the examiners had correctly identified the subject. Unverified cases were those where there was no knowledge of ground truth. The truthful versus deceptive category was based on polygraph examiner judgment.

The examinations used a standard-control question technique with some modifications. All examinations included, in order, a pretest interview, a control question test chart, a stimulation test chart, and another control question test chart. In 48 of the 112 examinations there were additional test charts. Examiners were to decide overall subject guilt or innocence, rate on a 6-point scale confidence in their judgment, and rate on a 5-point scale the ease of interpreting each physiological measure. Electrodermal activity, cardiovascular activity, and abdominal respiration were recorded. Twenty-six subjects also had thoracic respiration recorded.

Results

Of the 1120 truth-deception judgments, 707 (63.1%) were correct and 15 (.13%) were inconclusive. An analysis of variance with within factors of verification (verified-unverified), truthfulness (truthful-deceptive), and crime type (person-property), and between factor experience (high-low) was computed. The dependent measure was percentage of correct judgments by each examiner in each category. There was no significant effect for experience. There were significant effects for the other factors including a significant verification X truthfulness X crime type interaction. The following tables demonstrate the percentage distribution in each category:

Person Crime

<u>Examiner Judgment</u>		<u>Verified Cases</u>		<u>Unverified Cases</u>	
		<u>Deceptive</u>	<u>Truthful</u>	<u>Deceptive</u>	<u>Truthful</u>
<u>Examiner Judgment</u>	Deceptive	78.6	50.0	80.0	62.1
	Truthful	21.4	50.0	20.0	37.9

Property Crime

<u>Examiner Judgment</u>		<u>Verified Cases</u>		<u>Unverified Cases</u>	
		<u>Deceptive</u>	<u>Truthful</u>	<u>Deceptive</u>	<u>Truthful</u>
<u>Examiner Judgment</u>	Deceptive	75.7	47.9	64.3	33.6
	Truthful	24.3	52.1	35.7	66.4

Comparisons also were made between records with only the basic set of three charts and those with more than the basic set. 52.5% in the former category and 71.1% in the later category were correctly classified. Consistency among examiners was .89 for verified and .85 for unverified records based on Hoyt's intraclass reliability coefficient. Data on examiner confidence and ease of interpreting record sets is also reported.

It should be noted that, in this study, all cases were from criminal suspects. Thus, innocent subjects would have obtained some knowledge about the crime. Additionally, the initial polygraph examination was based on an overall assessment of guilt or innocence and any scoring criteria used were not reported.

Frank S. Horvath and John E. Reid. "The Reliability of Polygraph Examiner Diagnosis of Truth and Deception." Journal of Criminal Law, Criminology and Police Science 62(2) (1971): 276-281.

Procedure

Forty sets of polygraph records were selected from the files of John E. Reid and Associates. Twenty sets were from verified guilty subjects and 20 sets were from verified innocent subjects. The charts contained 164 relevant questions of which 81 were confirmed as deceptive answers and 83 were confirmed as truthful answers. The forty sets of records were from twenty-five cases involving theft, sexual misconduct, sabotage, bribery, and criminal damage to property. The polygraph records contained no sets that were very obvious and no sets that were considered uninterpretable. Thoracic and abdominal respiration, blood pressure-pulse rate, muscular movements, and electrodermal activity were recorded. Examinations used the standard Reid Control-Question Technique.

Ten Reid examiners reanalyzed the records. They were not familiar with the cases and they agreed not to discuss any of the charts among themselves while the research was in progress. Seven examiners had over a year of experience; three had only four to six months of experience and were still in their internship training program. The examiners did not know the number of truthful and deceptive sets of records. They were not allowed to call any set inconclusive, but they could call individual questions inconclusive.

Results

The ten examiners correctly classified 85% of the guilty cases and 90.5% of the innocent cases. Experienced and inexperienced examiners differed in percentage of correct classifications. The seven experienced examiners correctly classified from 85% to 97.5% of the cases, with an average of 91.4%. The three inexperienced examiners correctly classified from 77.5% to 90%, with an average of 79.2%.

Analysis of individual questions revealed that, for truthful questions, 82.7% were correctly classified, 9.6% were incorrectly classified, and 7.7% were inconclusive. For deceptive questions, 81.4% were correctly classified, 12.5% were incorrectly classified, and 6.2% were inconclusive. Omitting inconclusives, 89.6% of the truthful questions and 86.7% of the deceptive questions were correctly classified. Experienced examiners had higher correct classification rates for individual questions than inexperienced examiners. It should be noted that it is not known from the presented data whether examiners tend to misclassify the same people.

Fred L. Hunter and Philip Ash. "The Accuracy and Consistency of Polygraph Examiners' Diagnosis." Journal of Police Science and Administration 1(3) (1973): 370-375.

Procedure

Twenty cases from the files of John E. Reid & Associates were selected of which 10 were verified truthful and 10 were verified as untruthful. All the charts were of a control question technique, recorded on a four-channel polygraph (cardio, electrodermal, thoracic and abdominal respiration). The 20 sets of charts had 88 relevant questions of which 41 were verified as untruthful answers and 47 were verified as truthful.

Seven examiners from John E. Reid & Associates analyzed each record on two occasions separated by at least a three-month interval. Examiners did not know they were analyzing the same records on the two occasions. Examiners analyzed the records by making a subjective judgment of truth or deception based on comparison of control and relevant questions. Six examiners had more than one year of experience and one examiner had four and a half months of experience (was still an intern).

Results

Examiners correctly classified 86% of the truthful cases and 87% of the deceptive cases. The results were the same for the first analysis and the second analysis. Omitting inconclusives, 86% of the truthful cases and 88% of the deceptive cases (occasions combined) were correctly classified. Occasion-to-occasion agreement for the same examiner on the individual cases ranged from 75% to 90% (average 85%).

The seven examiners correctly classified 81% of the truthful questions and 83% of the deceptive questions (occasions combined). Omitting inconclusives, 85% of the truthful questions and 88% of the deceptive questions were correctly classified. Occasion-to-occasion agreement for the same examiner ranged from 67% to 91% (average 81%).

It should be noted that it is not known whether examiners tended to misclassify the same people, and that, in the analyses, each misclassification by each examiner is treated as an error. The authors note that the poorest results were obtained from the least-experienced examiner.

Verne W. Lyon, Deception Tests with Juvenile Delinquents. Journal of Genetic Psychology 48(3) (1936): 494-497

Procedure

One hundred cases were selected at random from the polygraph files of the Institute for Juvenile Research in Chicago. All of the polygraph charts were of juveniles of which 81 were boys, and 19 were girls, over age 15, who were suspected of crimes but had not confessed, and positive corroboration by other means had not yet been established.

The instrument made a record of respiration and circulatory activity. The technique was unique, but is in the general category of relevant-irrelevant. During the first recording no questions were asked. "During the second recording there were three indifferent questions, free from any complex situations, and so framed that the truth of the answers can be checked at once." The critical questions followed, uninterrupted by any indifferent stimuli. All questions were phrased to be answered "yes" or "no."

The institute attempted to follow up on these 100 cases to see if their results were corroborated by positive evidence or confession. There were no inconclusive cases among those reviewed.

Results

Of the 100 polygraph tests, 20 were called "clear", or truthful; 80 cases were reported as "disturbed records," that were "presumptive of deception." Of the 20 "clear" or truthful cases, only 7 (35 percent) could be followed up. All 7 cases were verified as innocent by factual evidence. Of the 80 "disturbed records" or deceptive cases, 33 (41 percent) of the subjects later confessed, many of them immediately after the test. Thus, the polygraph examiner correctly classified 100% of the verified cases.

Jesse Orlansky. An Assessment of Lie Detection Capability, Institute for Defense Analyses, July 31, 1962. Report UBG 62-641

Procedure

The United States Army Military Police prepared a report for the assessment in which they reported on the follow up of 3,153 cases in which the examiner reported "no deception indicated."

Results

There were 3,080 (97.7%) "instances in which results of examination in this category were in accord with results of other investigative techniques."

There were 73 (2.3%) "instances in this category in which contrary results were obtained through investigation or interrogation."

The extent to which there may be other errors is unknown. Those found nondeceptive, and not otherwise found guilty by court or appear guilty by contrary evidence, do not complain.

Robert B. Peters. "A Survey of Polygraph Evidence in Criminal Trials." The American Bar Association Journal 68 (February 1982): 162-165.

Procedure

The author, a polygraph examiner at the Wisconsin State Crime Laboratory, surveyed the results of stipulated polygraph examinations conducted at the state crime laboratory during the 33-month period from 1976 to 1979. Stipulation means that the prosecution, defense and judge agree to admit results of the polygraph examination into evidence, with the examiner testifying and subject to cross-examination. Actually, the examiner testified in only 11 of the cases.

Initially, there were 220 cases. Twenty-five were eliminated because they were inconclusive (not clearly indicative of truth or deception). Twenty-three cases were eliminated because no opinion could be given (e.g. subject unfit, made statements that resolved the issue of the test). The disposition of the 89 cases the examiner classified as truthful and the 83 cases classified as deceptive was determined.

Polygraph tests used standard control-question procedures. Respiration, electrodermal activity, and cardiovascular activity were recorded. Charts were numerically scored with a score of +6 or more considered truthful, -6 or less deceptive, and between -5 and +5 inconclusive.

Results

The following table shows the polygraph examiner's decision versus the disposition of the case (i.e. whether the outcome of the case reflected truthfulness or deception by the examinee):

		<u>Disposition of Case</u>	
		<u>Guilty</u>	<u>Innocent</u>
<u>Examiner's Decision</u>	<u>Deceptive</u>	62	12
	<u>Not Deceptive</u>	1	88

Eighty-five of the 89 cases classified truthful by the polygraph examiner involved criminal suspects. One of the suspects was convicted at trial and 84 suspects had charges dropped. Four of the truthful cases involved alleged victims or witnesses. In all 4 cases, the accused was convicted at trial making these subjects truthful.

Of the 83 cases classified as deceptive by the examiner, 2 claimed to be victims of crimes; the rest were criminal suspects. Of the suspects, 51 pleading guilty, 1 was acquitted, 3 had charges dropped with conditions (not included in table), 11 had charges dropped, 6 cases were pending (not included in the table), and 9 were convicted. In the cases involving the 2 alleged victims, charges against the accused were dropped at trial making these subjects deceptive.

The following table illustrates the outcome of cases that went to trial (the examiner testified in only 11 of the cases).

		<u>Examiner Testifying</u>		<u>Examiner Not Testifying</u>	
		<u>Convicted</u>	<u>Acquitted</u>	<u>Convicted</u>	<u>Acquitted</u>
<u>Examiner's</u>	<u>Deceptive</u>	9	1	9	0
<u>Decision</u>	<u>Truthful</u>	1	0	0	0

As a measure of validity, the Peters survey must be tempered by the realization that the results of the polygraph examination had a direct influence on those who made decisions to prosecute or drop charges. However, it was not the sole basis for those decisions as is evidenced by the decision to prosecute one who was called truthful. Polygraph results were also influential in bolstering the testimony of witnesses and victims in the view of those who decided to use them in court, despite the fact that the polygraph examiner did not testify in their behalf.

David C. Raskin. "Reliability of Chart Interpretation and Sources of Errors in Polygraph Examinations." Report No. 76-3 Contract 75-NI-99-0001. National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice, Department of Psychology, University of Utah, Salt Lake City, Utah, June 7, 1976.

Procedure 1

Sixteen sets of polygraph records from a previous study (Barland and Raskin, 1976) were used. Twelve were from confirmed guilty subjects and four were from confirmed innocent subjects. All examinations had used the federal modification of Backster zone-comparison technique consisting of three or more charts of ten questions each. Physiological responses were recorded on a field polygraph.

Twenty-five field polygraph examiners with different types of training and experience (18 with at least one year experience; 13 with formal training in numerical scoring) rescored the records. Seven examiners explicitly used numerical scoring.

Results

Of the 400 judgments, 78.8% were correct, 8.2% were incorrect, and 13% were inconclusive. Excluding inconclusives, there were 90.5% correct decisions. There were 33 errors - 20 false positives and 13 false negatives. (It is not possible from the presented data to determine the percentage correct for guilty and innocent subjects separately).

The number of correct decisions did not differ with length of examiner experience (more than one year versus less than one year). The ten examiners who had attended school emphasizing numerical scoring had higher accuracy than the fifteen examiners who attended other schools (97.1% versus 86.9%). The seven examiners explicitly using numerical scoring had a significantly higher accuracy than the eighteen who did not (98.9% versus 87.9%). Comparison among examiners trained in numerical scoring revealed that the seven examiners explicitly using numerical scoring had higher accuracy than the six not explicitly using numerical scoring (98.9% versus 88.5%).

Procedure 2 and Results

Different samples of polygraph records were analyzed to determine whether decisions of polygraph examiners depended on the person or agency involved in case referral. These analyses were computed to test the "friendly polygrapher" proposition of Orne (1975). Orne suggested that successful detection depends on motivation to deceive and threat of consequences. Thus, subjects referred for polygraph examinations by defense attorneys should be less easy to detect since they know results will not be used if they are found deceptive. The first sample consisted of 204 criminal cases from an experienced examiner who used control-question techniques. Ninety-eight cases were examined for defense counsel and 106 were examined for law enforcement authorities. Of the defense cases, 77.6% were classified truthful, 20.4% deceptive, and 2% inconclusive. Of the law enforcement cases, 75.5% were truthful, 19.8% deceptive, and 4.7% inconclusive. Thus, the two groups did not differ in type of decision.

A second sample (from Raskin and Barland, 1976) involved cases from two private polygraph firms. Nineteen examiners used control-question tests with criminal suspects. Fifty-four cases were cases for defense counsel and 57 were cases for law enforcement authorities or the subject's employer. Comparison of numerical scores for at least two charts of each set revealed no significant differences by type of case.

A third sample involved control-question tests conducted by the author on criminal suspects. Fourteen cases were from defense counsel and 13 were from law enforcement officials. Each case involved at least three charts and were numerically scored. Defense counsel cases were significantly more deceptive than law enforcement cases. Thus, there was no support for the "friendly polygrapher" hypothesis.

Procedure 3 and Results

Twelve cases involving conflicting polygraph examinations are analyzed to determine reasons for polygraph error. Descriptive data are presented.

Stanley M. Slowick and Joseph P. Buckley. "Relative Accuracy of Polygraph Examiner Diagnosis of Respiration, Blood Pressure, and GSR Recordings." Journal of Police Science and Administration 3 (3) (1975): 305-309.

Procedure

Polygraph charts from 15 truthful and 15 untruthful subjects whose cases had been independently verified and supported the original examiners conclusions were randomly selected from the laboratory case files of John E. Reid & Associates. The cases involved theft, industrial sabotage, drug abuse, and rape. Seven experienced staff examiners (average experience 3.8 years) participated in the study in which they were instructed to render decisions on the subject's veracity of 141 relevant questions (71 verified truthful and 70 verified untruthful) and 30 case issues by reviewing the complete charts. The examiners were not given the test questions or any other case information, and were instructed not to discuss their cases. Examiners also made decisions about the 30 cases and 141 relevant questions analyzing each physiological measure independently. This was done by analyzing each physiological measure on a different occasion separated by three month intervals and masking two channels while the third was analyzed. Abdominal or thoracic respiration, blood pressure/pulse rate, and electrodermal activity were the physiological measures.

Results

Chart review as a composite of three physiological measures demonstrated that the examiners correctly classified 91% of the truthful cases and 84% of the untruthful cases. Omitting inconclusives, 93% of the truthful and 85% of the untruthful cases were correctly classified. The individual question analysis revealed that 88% of the truthful questions and 73% of the untruthful questions were correctly classified. Omitting inconclusives, these figures are 93% and 80%, respectively.

The analysis by independent physiological measure at three month intervals revealed slightly lower correct classification rates. Differences between physiological measures in overall classification rates were small. It should be noted that it is not possible to determine from the presented data if examiners tended to misclassify the same subjects. The presented data consider each misclassification by each examiner an error.

Douglas E. Wicklander and Fred L. Hunter. The Influence of Auxiliary Sources of Information in Polygraph Diagnosis." Journal of Police Science and Administration 3(4) (1975): 405-409.

Procedure

Twenty sets of polygraph records selected from the files of John E. Reid & Associates, of which 10 sets were those of verified truthful subjects and 10 sets of verified untruthful subjects. The sets contained 89 relevant questions, 43 verified deceptive and 46 verified truthful. The instruments used were four channel Stoelting polygraphs (recording thoracic and abdominal respiration, blood pressure-pulse rate, and electrodermal activity) and the technique was Reid Control-Question Technique. Six experienced polygraph examiners on the Reid staff (average experience = 4 years) read each set of polygraph records on two different occasions, with a two-month lapse to eliminate recall. Each examiner had to call each set and each question truthful or deceptive without discussing his opinions with others. The judgment was subjective based on comparison of relevant and control questions.

Prior to the first review of charts the examiner was told only the type of case involved. Prior to the second review of the charts the examiner was given a synopsis of the case history, subject data, verbal and nonverbal behavior symptoms routinely noted by the original examiner, and the relevant test questions. Of interest was whether this extra information would influence chart analysis.

Results

Examiners correctly classified 87% of the truthful cases and 90% of the deceptive cases on the first chart review. Omitting inconclusives, 95% of the truthful and 92% of the deceptive cases were correctly classified.

At the second review 87% of the truthful cases and 98% of the deceptive cases were correctly classified. Omitting inconclusives these figures are 91% and 98%, respectively. There were fewer inconclusive judgments at the second review.

On the first review of the individual question analysis, 80% of the truthful questions and 86% of the deceptive questions were correctly classified (omitting inconclusives, 91% and 91%, respectively). On the second review, 82% of the truthful questions and 95% of the deceptive questions were correctly classified (omitting inconclusives, 92% and 97% respectively). It should be noted that it cannot be determined from presented data whether examiners tended to misclassify the same people. Each misclassification is treated as an error.

Jan Widacki. "Analiza Przeszanek Diagnozowania W. Badanich Poligraficznych." Uniwersytetu Slaskiego, Katowice, 1982 ("The Analysis of Diagnostic Premises in Polygraph Examinations.")

Procedure

All of the polygraph cases conducted on criminal suspects by the Department of Criminalistics of the Silesian University (Poland) in the years 1978-1979 which were verified by subsequent legal proceedings were analyzed. Sixteen people were judged guilty by both the polygraph examiner and subsequent legal proceedings. Twenty-two subjects were judged innocent by the examiner and subsequent legal proceedings.

Results

Control question tests were analyzed by the Backster numerical scoring method. In 91.6% of the cases, guilty and innocent subjects were differentiated (separate percentages are not given for guilty and innocent).

With peak of tension tests, 80% of the guilty persons reacted to the relevant questions. The authors say that 60% of the innocent subjects did not react to the relevant questions and 40% showed only minor reactions to relevant questions. Visual biofeedback on peak of tension tests enhanced the ability to discriminate guilty subjects from innocent subjects. It should be noted that the authors give incomplete details about a number of aspects of the study.

Laboratory Abstracts

Control Question Techniques

Gordon H. Barland (Prepared By). "A Validity and Reliability Study of Counterintelligence Screening Test." Security Support Battalion, 902d Military Intelligence Group, Fort George G. Meade, Maryland, May 12, 1981.

Procedure

Fifty-six U.S. Army Employees (38 military, 18 civilian; 40 men, 16 women; age range 21 to 55 years; educational level 12 to 17 years) were subjects in a mock screening situation to determine polygraph accuracy with the Counterintelligence Screening Test with directed lie questions. This test is derived from the federal version of the zone comparison test. The test contains 13 questions - five relevant, four directed lies, and two symptomatic. Polygraph tests were administered by three polygraph examiners trained at the U.S. Army Military Police School polygraph course, certified by the Department of the Army with three, six, and nine years of polygraph experience. Procedures were used to standardize testing procedure among the examiners. Five different models of field polygraphs were used. Respiration, skin resistance, and relative blood pressure were recorded.

Questions involved subject's data of birth, place of birth, education, residences, and employment. All subjects initially filled out a biographical data sheet with truthful information. The information was confirmed by background investigations of each subject. Subjects then filled out new biographical data sheets (some giving false information) to give to the examiner. It was randomly determined which subjects would answer all questions truthfully and which would lie to one question. The question subjects lied about varied. Deceptive subjects were told they would receive \$20 if they appeared truthful on the polygraph test. The rationale of lie detection was explained to all subjects. After a pretest interview, subjects received a stimulus test. The biographical data sheet was then reviewed. The polygraph test consisted of a minimum of three presentations of 13 questions. If the examiner was uncertain of the subject's status at the end of three charts, additional charts were run with a maximum of six. If the examiner was still uncertain after six charts, the subject was called inconclusive. The examiner was not allowed to interrogate the subject.

Three methods were used to evaluate the charts:

1. Zone method - Each relevant question was evaluated against the larger control question in its zone for each channel and scored from +3 to -3 as taught at the U.S. Army Police School Polygraph course. Scores for each relevant question were summed across charts. The answer to a question was designated as truthful if the score to a question was +3 or greater, deception if the score was -3 or lower, and inconclusive if the score was between +3 and -3.

2. Greatest control - The method used the same procedure as the zone method except that the five relevant questions were evaluated against the one control question on the chart with the largest overall reaction.

3. Relevant-irrelevant method - There was no numerical scoring and each relevant question was evaluated in terms of size and consistency of response without reference to control questions. For all methods, if the largest response on the chart occurred to one of the symptomatic questions, the subject was classified inconclusive.

Results

The first set of analyses used as data the examiners' classification of the subjects as deceptive or not deceptive irrespective of whether the examiner had identified accurately the deceptive question. The following results were obtained with the zone method:

		<u>Actual</u>	
		<u>Deception</u>	<u>No Deception</u>
<u>Predicted</u>	<u>Deception</u>	21	5
	<u>No Deception</u>	5	16
	<u>Inconclusive</u>	4	5

Omitting inconclusives, 81% of the guilty and 76% of the innocent were correctly classified.

The following results were obtained with the greatest control method:

		<u>Actual</u>	
		<u>Deception</u>	<u>No Deception</u>
<u>Predicted</u>	<u>Deception</u>	15	4
	<u>No Deception</u>	7	20
	<u>Inconclusive</u>	8	2

Omitting inconclusives, 68% of the guilty and 83% of the innocent were correctly classified.

The following results were obtained with the relevant/irrelevant method:

		<u>Actual</u>	
		<u>Deception</u>	<u>No Deception</u>
<u>Predicted</u>	<u>Deception</u>	24	6
	<u>No Deception</u>	4	19
	<u>Inconclusive</u>	2	1

Omitting inconclusives, 86% of the guilty and 76% of the innocent were correctly classified. All three methods were able to identify nondeceptive subjects significantly greater than chance. Zone comparison and relevant/irrelevant were able to identify non-deceptive subjects greater than chance.

Results also were analyzed based on individual questions. There were 250 truthful questions and 30 deceptive questions. Nondeceptive questions were identified significantly greater than chance by all methods (range 91% to 97% correct classifications, omitting inconclusives). The relevant/irrelevant method

was the only method able to identify deceptive questions significantly greater than chance (range 54% to 69% correct classifications, omitting inconclusives). Data also are provided on the stimulus test. Subjects received more than one chart on this test if the examiner's first or second choice was not the one selected by the subject. Classification accuracy (global assessment of deception/no deception) was significantly related to number of charts on the stimulus test.

Gordon H. Barland and David C. Raskin. "An Evaluation of Field Techniques in Detection of Deception." Psychophysiology, 12(3)(1975): 321-330.

Procedure

Seventy-two college students (sex not specified) were subjects in an experiment to evaluate detection of deception field techniques. Subjects were randomly assigned to guilty and innocent conditions. Guilty subjects were required to steal \$10 from a desk drawer and could keep the money if they deceived the examiner. Innocent subjects were told the details of the crime and that failure to appear innocent reflected lack of cooperation and they may not get course credit for participating.

After a pretest interview involving collection of background information explanation of the rationale of lie detection, and at least two reviews of the questions on the test, subjects received three presentations of ten questions. A control - question technique (the Federal Government modification of the zone comparison polygraph test) was used. Between charts one and two, a card test was administered with some subjects getting positive feedback, others negative feedback, and others no feedback. The author noted that this manipulation had no effect and did not present the results. The polygraph examiner (experience not discussed), blind to subject status, scored the tests immediately by semi-objective scoring using the U.S. Army Military Police School criteria. Each of the control/relevant question pairs was scored from -3 to +3. A score of +5 or greater was considered truthful, -5 or lower guilty, and between +4 inconclusive. The polygraph records also were scored by five additional examiners, blind to subject status with intensive experience who had graduated from the U.S. Army Military Police School. An objective quantitative scoring also was conducted by people unaware of the study and unfamiliar with field techniques using more physiological measures.

Cardiovascular activity, skin resistance, and thoracic or abdominal respiration were recorded with a field model polygraph. The authors note that the recording of skin resistance and cardiovascular activity were not of the quality obtained from newer field polygraphs.

Results

The following results were obtained with the semi-objective scoring procedure:

		<u>Guilty</u>	<u>Actual</u>	<u>Innocent</u>
<u>Predicted</u>	<u>Guilty</u>	23		6
	<u>Innocent</u>	3		15

Of the seventy-two subjects, 53% were correctly classified, 12% incorrectly classified, and 35% inconclusives. (The inconclusives are omitted from the above table because the authors did not report their actual status). Omitting inconclusives, 81% were correctly categorized. Analysis of variance on the semi-objective scores indicated that all physiological measures discriminated between guilt and innocence but skin resistance was superior.

There was considerable agreement between the five additional examiners and the original examiner. Percentage of correct classifications ranged from 79% to 86% (mean = 81.7%) for the six examiners. Thus, the blind examiners were similar to the original examiner. Correlations between pairing of examiners for total semi-objective numerical scores ranged from .78 to .95 (mean = .86). Of the 1,080 pairwise comparison, the examiners agreed on 66.3%. There were only 25 instances (2.3%) where the examiners reached opposite conclusions.

Data on the effect of changing the inconclusive range and the objective quantitative analysis also are provided.

M. T. Bradley and Michel Pierre Janisse. "Accuracy Demonstrations, Threat, and the Detection of Deception: Cardiovascular, Electrodermal, and Pupillary Measures." *Psychophysiology* (1981) 18(3): 307-315.

Procedure

One hundred ninety-two male college students were subjects in a study to determine the effect of several variables on detection of deception. Guilty subjects participated in a mock crime involving stealing and hiding an envelope containing \$1.00. All subjects were told that they were accused of the crime and were to deny involvement. One-half of the innocent and one-half of the guilty subjects were told they would receive a painful electric shock (the shock was not administered) if they were judged guilty.

Pupil size, heart rate, and skin resistance were recorded. The polygraph examiner was blind to subject status. The experience of the polygraph examiner is not discussed. Subjects were given both a control-question test and a guilty knowledge test about the crime (tests were counterbalanced). Prior to the polygraph test, the questions on the control-question test were reviewed with the subject. Subjects were allowed to discuss the questions and eliminate ambiguities about the control questions. The control-question test consisted of nine questions (three relevant pairs) presented three times. Questions on the guilty knowledge test were not reviewed with the subject prior to the polygraph test. This test consisted of four, five-item question sequences plus a buffer question for each sequence; the test was presented once to the subject.

Before the polygraph test, a series of rigged card tests with three trials were tested. Subjects were given false feedback about the number of times the card they had selected had been detected. Subjects were falsely told that from 0 to 3 cards had been detected.

Each control-relevant pair from the control-question test was scored 1, 0, or -1 and summed across questions. A score of +2 or greater reflected innocence, -2 or less reflected guilt; scores between +2 and -2 were classified as inconclusive. On the guilty knowledge test, responses to each of the four sequences were scored 2, 1, or 0 based on the relative magnitude of the response to the critical item.

Results

The following table demonstrates frequency of detection for each physiological measure on each test:

	<u>Actual</u>			
	<u>Skin Resistance</u>			
	<u>Control Question Test</u>		<u>Guilty Knowledge Test</u>	
	<u>Guilty</u>	<u>Innocent</u>	<u>Guilty</u>	<u>Innocent</u>
<u>Guilty</u>	58	9	57	11
<u>Innocent</u>	13	56	39	85
<u>Inconclusive</u>	25	31		
<u>Predicted</u>	<u>Pupil Size Change</u>			
<u>Guilty</u>	33	30	32	19
<u>Innocent</u>	21	25	64	77
<u>Inconclusive</u>	42	41		
	<u>Heart Rate Change</u>			
<u>Guilty</u>	34	19	43	17
<u>Innocent</u>	20	32	53	79
<u>Inconclusive</u>	42	45		

Skin resistance was the most efficacious measure in separating guilty and innocent people. With skin resistance, 82% of the guilty subjects and 86% of the innocent subjects were correctly detected (excluding inconclusives) on the control-question test. With the guilty knowledge test, 59% of the guilty subjects and 89% of the innocent subjects were correctly classified using skin resistance. Detection accuracy results for a combination of measures are not presented.

Analysis of variance on the numerical scores demonstrated that, for the control-question test, punishment influenced the change in heart rate measure and manipulated effectiveness was directly related to detection detectability with skin resistance. In the guilty knowledge test, punishment did not affect any of the dependent measures, but there were significant effects for manipulated effectiveness with skin resistance.

The authors suggest that the detection accuracy with the control-question test and guilty knowledge test should not be directly compared because they used a weak version of the guilty knowledge test (few items of information) and they used less than optimal procedures for the control-question test. That is, for the control-question test, they did not repeat trials for inconclusive subjects, they did not combine physiological measures for detection and their semi-objective scoring system had a limited range.

Robert J. Gatchel, John E. Smith, and Norman M. Kaplan. "The Effect of Propranolol on Polygraphic Detection of Deception." (1983) Unpublished Manuscript, University of Texas Health Sciences Center.

Procedure

Twenty-eight males aged 20 to 40 years were paid \$15 for participating in a study to determine the effect of propranolol on detection of deception in a mock crime situation. Subjects were randomly assigned to one of four groups (seven subjects per group); (1) Guilty - propranolol, (2) Guilty - no propranolol; (3) Innocent - propranolol, and (4) Innocent - no propranolol. The guilty subjects were required to steal and then hide a piece of jewelry from an office in one hour. Innocent subjects were told that a mock crime had taken place but were not told details. All subjects were told to deny involvement in the crime. Both innocent and guilty subjects could receive a \$10 bonus for appearing truthful on the test. Drug subject received a 40 mg. tablet of propranolol one to two hours before the polygraph test.

Blood pressure, electrodermal activity, and abdominal and thoracic respiration were recorded. Two professional polygraph examiners (each testing 14 subjects), blind to subject's assigned group status, administered a nine-item control question polygraph test. Detection was based on pairs of control-relevant questions (the last six questions). Each pair was scored (by the examiner) from +3 to -3 based on the size of the relative difference. Respiration was scored based on the clearest channel. Scores were summed across pairs and channels. Subjects with score +6 or higher were designated truthful, -6 or lower were designated as deceptive, and inconclusive between +5. Examiners also rated on a seven-point scale whether they thought the subject had taken a drug.

Results

Results are listed in the following table:

		<u>Actual</u>			
		<u>Guilty:</u>	<u>Innocent:</u>	<u>Guilty:</u>	<u>Innocent:</u>
		<u>No Drug</u>	<u>No Drug</u>	<u>Drug</u>	<u>Drug</u>
Predicted	Guilty	3	0	4	0
	Innocent	1	4	0	7
	Inconclusive	3	3	3	0

Analysis of variance revealed a significant guilt-innocence effect, but no significant drug effects or interactions. Examiners were correct about subject's drug status in 71% of the cases; however, they were not very confident about their ratings. The authors note that their study has more inconclusives than other studies. They suggest that this reflects conservative behavior on the part of the examiners since the examiner knew that drugs were being administered.

David L. Hammond. "The Responding of Normals, Alcoholics, and Psychopaths in a Laboratory Lie-Detection Experiment." Unpublished Doctoral Dissertation, California School of Professional Psychology, 1980.

Procedure

Normal college students (11 innocent and 10 guilty), alcoholics (10 innocent and 10 guilty) and psychopaths (9 innocent and 12 guilty) participated in the study. All subjects were male, aged 21 to 55 years, and paid \$7.00 for participating. Guilty subjects participated in a mock crime involving thefts of \$10.00 and were told not to confess. Instructions were presented by tape and were designed to motivate the subjects. Innocent subjects listened to the same tape and, thus, knew the general nature of the crime. All subjects could receive an additional \$10.00 for a truthful outcome on the test. Alcoholics had previous treatment for alcoholism and met certain diagnostic criterion. Psychopaths had a prior prison record and met other diagnostic criteria. Respiration, electrodermal activity, and cardiovascular activity were recorded. Inexperienced examiners from the Backster School of Lie Detection in the fifth and sixth weeks of their seven-week training session conducted the tests. The Backster "You Phase" control-question procedure was used. The test was scored with the zone-comparison method. Testing involved a minimum of three charts, but in most cases two charts were used for analysis. Numerical scores greater than 8 were considered truthful, less than 8 were considered deceptive, and between +8 and -8 were inconclusive. An expert examiner also independently evaluated the polygraph charts.

Results

The following results were obtained for the student examiners:

		<u>Actual</u>	
		<u>Guilty</u>	<u>Innocent</u>
<u>Predicted By Student Examiners</u>	<u>Guilty</u>		
	Normal	6	3
	Alcoholic	9	1
	Psychopath	8	2
	<u>Innocent</u>		
	Normal	1	4
	Alcoholic	0	5
	Psychopath	0	3
	<u>Inconclusive</u>		
Normal	3	4	
Alcoholic	1	4	
Psychopath	4	4	

Omitting inconclusives, 86% of the normal guilty, 100% of the alcoholic guilty, and 100% of the psychopathic guilty were correctly classified by the student examiners. Omitting inconclusives, 57% of the normal innocent, 83% of the alcoholic innocent, and 60% of the psychopathic innocent were correctly classified.

The following results were obtained for the expert examiner:

		<u>Actual</u>	
		<u>Guilty</u>	<u>Innocent</u>
<u>Predicted</u> <u>By Expert</u> <u>Examiner</u>	<u>Guilty</u>		
	Normal	4	2
	Alcoholic	8	0
	Psychopath	9	0
	<u>Innocent</u>		
	Normal	0	0
	Alcoholic	0	2
	Psychopath	0	2
	<u>Inconclusive</u>		
Normal	6	9	
Alcoholic	2	8	
Psychopath	3	7	

Omitting inconclusives, 100% of the guilty subjects in each condition were correctly classified by the expert examiner. Omitting inconclusives, the expert examiner correctly classified 0% of the normal innocents, and 100% of the alcoholic and psychopathic innocents. Data also are presented using other inconclusive ranges.

Analysis of variance on the numerical scores revealed no significant effects for examiner experience, personality, or physiological measures. Student examiners' numerical scores and expert examiner's numerical scores were correlated .64 for the innocent condition and .61 for the guilty condition.

Charles Robert Honts. "The Effects of Simple Physical Countermeasures on the Physiological Detection of Deception." Unpublished Master's Thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, May, 1982. [Also presented at the Society for Psychophysiological Research, Minneapolis, Minnesota, October 1982, with Robert L. Hodes.]

Procedure

Forty-eight students (males and females, mean age 18.8 years) were participants in a study to determine the effect of physical countermeasures on the detection of deception, using a control question test. Subject's were randomly assigned to one of four groups (12 subjects per group); (1) Innocent, (2) Guilty Controls - no countermeasure, (3) Guilty - pair countermeasure, and (4) Guilty - muscle tension countermeasure. All guilty subjects participated in a mock crime which involved stealing an examination. The guilty control subjects were told that it may be possible to deceive the polygraph examiner. Guilty-pain countermeasure subjects were given training on biting the tongue at particular points during the control question test and were told to relax during relevant questions. Guilty-muscle tension subjects received training on pressing their toes against the floor and tensing their muscles at critical points during the test and also were told to relax during relevant questions. Innocent subjects knew the general nature of the mock crime, but were naive to its details. They were told that if they were honest the examiner would find them truthful. All subjects were told they would get extra points applied to their final course grades if they produced a truthful outcome on the polygraph.

All subjects received two experimental sessions. During the first session, subjects were assigned to a group and received training if they were to use countermeasures. The second session (within one week of the first) was the polygraph exam. All guilty subjects were encouraged to appear innocent to receive bonus points. Subjects were cautioned not to be detected using countermeasures or they would lose bonus points.

Thoracic or abdominal respiration, skin resistance, and cardiovascular activity were recorded. A licensed examiner with five years of field experience trained at the Backster School of Lie Detection, administered the polygraph tests. The examiner did not know the guilt/innocence status of the subjects, but did know the proportions of guilt and innocent subjects and did know that muscle or pair countermeasures would be used. Subjects received four presentations of eight questions.

Immediately after the polygraph test, the examiner was required to make an innocent/guilt/inconclusive decision based on limited chart analysis (less than five minutes) and to judge whether countermeasures were used. Two weeks after the last subject was tested, the examiner scored the charts by semi-objective methods and judged whether countermeasures were used from the physiological data. Data were sent to an experienced evaluator for a blind evaluation using the semi-objective method. An objective, quantitative analysis of the data also occurred. The semi-objective analysis used the Backster Spot Analysis Technique in which pairs of responses were rated from -3 to +3 (based on the relative magnitude of relevant and associated control question) and summed. Post examination debriefings indicated that five subjects in the countermeasure groups failed to employ the designated

countermeasures and three subjects in the guilty-control group tried to use countermeasures. Data were analyzed with regard to subject's debriefing reports.

Results

The original examiner was not able to detect countermeasure usage either from a subject's behavior or from the polygraph charts. The blind rescorer also could not detect countermeasure usage.

The ability of the original examiner to correctly classify subjects based on a limited chart analysis is reflected in the following table (numbers in parentheses are those subjects that did not comply with the group's instruction).

		<u>Actual</u>			
		<u>Guilty: Control</u>	<u>Innocent</u>	<u>Guilty: Pain</u>	<u>Guilty: Muscle</u>
<u>Predicted</u>	<u>Guilty</u>	5(1)	4	5(3)	7(2)
	<u>Innocent</u>	1	6	1	1
	<u>Inconclusive</u>	3(2)	2	3	2

Results for semi-objective analysis (+12 or greater indicating innocence, -12 or lower deception, and between +12 and -12 inconclusive based on the three most productive charts) are in the following table:

		<u>Actual</u>			
		<u>Guilty: Control</u>	<u>Innocent</u>	<u>Guilty: Pain</u>	<u>Guilty: Muscle</u>
<u>Predicted</u>	<u>Guilty</u>	6(1)	2	5(3)	6
	<u>Innocent</u>	0	4	1	0
	<u>Inconclusive</u>	3(2)	6	3	4(2)

Excluding inconclusives, the examiner was correct 89% of the time. Countermeasures did not affect the results of either the limited chart analysis or the semi-objective analysis. Analysis of variance for semi-objective scores from four charts (with the five noncompliant subjects in the countermeasure groups removed) also did not reveal a countermeasure effect. Objective quantitative analysis of each physiological measure found significant countermeasure effects for finger pulse amplitude, the results for respiration amplitude approached significance.

Semi-objective scores of the original examiner were significantly correlated ($r = .88$) with those of the blind evaluator. When there was a decision of truthful or deception, there was 100% agreement between the raters.

Raskin, D. C., and Hare, R. D. Psychopathy and Detection of Deception In a Prison Population. Psychophysiology, 15 (1978); 126-136

Procedure

Twenty-four psychopaths (mean age = 23.1 years) and 24 non-psychopaths (mean age = 26.5 years), from an inmate population, participated in the study. Twelve psychopaths and 12 non-psychopaths were innocent and 12 psychopaths and 12 non-psychopaths were guilty of a mock crime involving the theft of \$20 from an adjacent room. Subjects were then given a polygraph test about the crime and could keep the \$20 if their test performance was successful. Guilty subjects were to deny involvement in the crime; innocent subjects were told about the crime and instructed to tell the truth. Psychopaths were selected by an experienced researcher based on an assessment of their behavior over a long period. The conception of psychopathy was consistent with that of Cleckley (1964).

Subjects were administered a control question polygraph test by an examiner trained in field polygraph techniques blind to subject status. Control questions were developed from a pretest interview. Relevant questions were constant for all subjects; control questions differed. Prior to the 10 questions, subjects were administered a stimulation test to demonstrate technique effectiveness. The polygraph test consisted of three charts; if results were not obvious, additional charts were run to a maximum of seven. In such situations, the fourth chart used a silent answer procedure. The examiner continually stressed the importance of being truthful and directed the attention of innocent subjects to control questions and guilty subjects to relevant questions.

Thoracic and abdominal respiration, skin conductance, skin potential, heart rate, finger blood volume, and finger pulse amplitude were recorded. Data were evaluated in two ways. The first was by the examiner using responses previous research had demonstrated significantly discriminated between guilt and innocence.

Each control and relevant question pair were compared, scored from +3 to -3, and summed; +6 or higher indicated innocence; -6 or lower indicated guilt. Results using all charts for each subject are presented in the following table:

		<u>Actual</u>	
		<u>Guilty</u>	<u>Innocent</u>
<u>Examiner</u> <u>Decision</u>	<u>Guilty</u> - (<u>Psychopath</u>)	12	1
	(<u>Nonpsychopath</u>)	9	1
	<u>Innocent</u> - (<u>Psychopath</u>)	0	11
	(<u>Nonpsychopath</u>)	0	10
	<u>Inconclusive</u> - (<u>Psychopath</u>)	0	0
	(<u>Nonpsychopath</u>)	3	1

Thus, 88% were correctly classified, 4% incorrectly classified, and 8% were inconclusive. Omitting inconclusives 96% were correctly classified. Psychopaths and nonpsychopaths did not significantly differ in classification accuracy. Comparison of classification based only on three charts per subject with classification using additional charts when necessary demonstrated a reduction in inconclusives with more charts, but similar correct classification rates. Optimal boundary was suggested to be in the range of +2 to +4. Analysis of variance and post-hoc comparisons on numerical scores revealed that skin conductance was the best guilt/innocence discriminator and the cardiovascular measures the poorest,

although all measures significantly discriminated between the groups. Respiration was better at identifying innocent subjects than guilty subjects. Effects of psychopathy on the individual measures were minimal. Results of a more quantitative analysis using additional physiological measures scored by a technician with no knowledge of the experiment also are reported. Characteristics of individual physiological measures associated with guilt and innocence are discussed.

L. I. Rovner, D. C. Raskin and J. C. Kircher. "Effects of Information and Practice on Detection of Deception." Presented at the Society for Psychophysiological Research, Madison, 1978.

Procedure

Seventy-two males participated in a study to determine the effect of information and practice on detection of deception. The 36 guilty subjects participated in a mock crime (stealing a ring). The 36 innocent subjects were told a theft had occurred. Prior to their polygraph test, subjects were divided into the following groups with 12 innocent and 12 guilty subjects in each group: (1) STD - Subjects merely waited in a room before their polygraph test, (2) INFO - Subjects were given information about the nature of the polygraph and how to appear innocent, and (3) INFO + Practice - Subjects were given the same information as group two, and two polygraph practice sessions with feedback about their performance. Examiners, (not professional polygraph examiners), blind to a subject's guilt or innocence, administered a control question polygraph test. Subjects were paid \$7.50 for participating in the study and \$10 if the outcome of their test was truthful. All subjects were to deny involvement in the theft. Skin conductance, blood pressure, respiration, and digital vasomotor activity were recorded; there were three charts per subject. Responses were scored with a numerical scoring system and an inconclusive range of ± 5 . Results are based on the scoring of a second examiner blind to test results and subject status.

Results

In both the STD and INFO groups, 88% were correctly classified, 4% were incorrectly classified, and 8% were inconclusive. Omitting inconclusives 95% were correctly classified. The one incorrect classification in the STD group and the 1 incorrect classification in the INFO group were false positives. The actual guilt or innocence of inconclusive subjects was not reported. INFO + Practice reduced the number of correct classifications and increased the number of false positives and negatives. That is, 62.5% were correctly classified, 25% were incorrectly classified (three false positives and three false negatives), and 12.5% were inconclusive. Analysis of variance and post-hoc comparisons revealed that the innocent subjects in the INFO + Practice group had significantly lower scores than the other two groups, whereas the three guilty groups did not significantly differ. The authors note that the lack of time and similarity between practice and the actual polygraph test may be important to observation of the effect.

William M. Waid, Emily Carota Orne, and Martin T. Orne. "Selective Memory for Social Information, Alertness, and Physiological Arousal in the Detection of Deception." Journal of Applied Psychology (1981) 66(2): 224-232.

Procedure

Thirty-four designated innocent and 40 designated guilty subjects (males—mean age 18 to 28 years) participated in a study to determine the relationship between detection of deception and memory of questions and words involved in the polygraph test. Guilty subjects overlearned 6 code words during timed interpolated tasks. Innocent subjects performed the same tasks but learned no words. Innocent subjects were told it was often difficult to prove one's innocence on a polygraph test. Guilty subjects were told that highly mature and intelligent people could deceive the examiner. Care was taken to involve innocent subjects in the procedures. Subjects were tested by a professional examiner blind to subject status; skin conductance, blood pressure, and respiration were recorded. Subjects were administered three guilty person tests and a guilty knowledge test. The guilty person tests consisted of 9 relevant questions, 8 control questions, and 5 irrelevant questions. A card test was administered between the first and second tests and subjects were questioned about their answers. In the guilty knowledge test, subjects were asked if each of 24 words had special meaning. The list was presented twice. Subjects were then asked to write down as many words as they could remember from the guilty knowledge test and as many questions as possible from the guilty person test.

Results

Different methods of quantifying the data were used to determine detection accuracy for the guilty person tests. One method averaged response to relevant and control questions. A subject was considered deceptive if average response to the relevant questions was larger than average response to control questions. The following results were obtained when skin conductance was used.

		<u>Guilty</u>	<u>Actual</u>	<u>Innocent</u>
<u>Predicted</u>	<u>Guilty</u>	29		8
	<u>Innocent</u>	11		26

72.5% of the guilty subjects and 76.5% of the innocent subjects were correctly classified.

When the detection criterion was showing a larger skin conductance response to relevant than control questions on 7 or more questions, the following results were obtained:

		<u>Guilty</u>	<u>Actual</u>	<u>Innocent</u>
<u>Predicted</u>	<u>Guilty</u>	22		8
	<u>Innocent</u>	18		26

Fifty-five per cent of the guilty subjects and 76.5% of the innocent subjects were correctly detected. Detection accuracy was much lower with respiration and with blood pressure.

There were results for memory of test items. Analysis of variance indicated that there were larger skin conductance responses to relevant and control questions that were subsequently recalled than to relevant and control questions not subsequently recalled on the first and second guilty person tests. On the third guilty person test subsequently recalled control questions elicited larger responses than control questions not subsequently recalled, but this was not true for relevant questions. On certain tests, more guilty subjects were called deceptive when skin conductance responses were larger to relevant than control questions, and more innocent subjects were called not deceptive when responses were larger to control questions than to relevant questions.

There was less of a relationship between recall of items and detection on the guilty knowledge test than with the guilty person test. There were no relationships between recall and blood pressure or respiration. Detection accuracy with the guilty knowledge test was not reported.

William M. Waid, Martin T. Orne, and Stuart K. Wilson. "Effects of Level of Socialization on Electrodermal Detection of Deception." Psychophysiology 16 (1) (1979): 15-22.

Procedure

Thirty male college students (aged 18 to 28 years) participated in an experiment to determine the effect of socialization on detection of deception. The fifteen subjects designated guilty overlearned 6 "code words" during a one hour procedure of time interpolated tasks. The fifteen innocent subjects performed the same tasks but learned no "code words." The experimenters emphasize that care was taken to involve both the innocent and guilty subjects in the task. Guilty subjects were told that highly intelligent, mature individuals could escape detection.

A professional polygraph examiner with no knowledge of a subject's status or the code words administered several types of polygraph tests. The six "code words" differed for the subjects to prevent bias in the examiner. Skin conductance was recorded. The following tests were administered (always in the same order) after a pretest interview to obtain personal history and a review of the relevant questions on the first test: a) Guilty Person Test - Subject was asked three questions about his guilt or innocence and two interpolated control questions. The subject was questioned about his answers, presented a stimulus test, and then a second Guilty Person Test. The first and second Guilty Person Tests differed in that the former used unreviewed control questions, whereas the later reviewed control questions with the subject; b) Peak-of-Tension Test - Subjects were asked how many correct code words existed. Two series of numbers were presented one with the sequence in ascending order and one with the sequence in descending order; c) Guilty Knowledge Test - Subjects were asked if each of 24 words had special meaning to them. Six of the words were code words. This list was repeated four times.

Subject then completed the socialization scale of the California Psychological Inventory and other questionnaires. Criteria for detection differed with type of test. On the Guilty Person Tests, subjects were classified as deceptive if the response to any of the three critical questions was larger than the largest response to the control questions. Deception was indicated on Peak-of-Tension if the largest response of the series was to the critical number (omitting the first number). For the Guilty Knowledge test a code word was considered detected if it elicited a larger response than the three code words in the same category. Subjects having code words detected seven or more times were considered deceptive.

Results

Results are in the following table:

		<u>Actual</u>	
		<u>Guilty</u>	<u>Innocent</u>
<u>Guilty Person 1</u>			
	Guilty	12	3
	Innocent	3	12
<u>Predicted</u>	<u>Guilty Person 2</u>		
	Guilty	12	3
	Innocent	3	12
	<u>Peak of Tension</u>		
	Guilty	7	3
	Innocent	8	12
	<u>Guilty Knowledge</u>		
	Guilty	8	1
	Innocent	7	14

There were significant effects for socialization. Deceptive subjects not detected scored significantly lower on the socialization scale than detected deceives for most of the tests. Among guilty subjects socialization was significantly positively correlated with conductance response magnitude (mean of four tests). Misclassified innocents had higher socialization scores than other innocents. Conductance response magnitude was significantly positively correlated with socialization in innocent subjects. It should be noted that the effectiveness of the different polygraph tests can not be determined in this study, since the tests were administered in the same order for all subjects.

Jan Widacki and Frank Horvath. "An Experimental Investigation of the Relative Validity and Utility of the Polygraph Technique and Three Other Common Methods of Criminal Identification." Journal of Forensic Science 23 (3) (July 1978): 596-601.

Procedure

Subjects were 80 student volunteers (42 males, 38 females; aged 19 to 24 years) at Jagollian University (Poland). Each subject provided the experimenter with hand-writing specimens, fingerprints and full-face photographs. Subjects were divided into 20 groups of four each. In each group there were three innocent persons and one perpetrator. Each perpetrator was involved in an independent, but similar case. The perpetrator was instructed to take an envelope to a specific building and give the envelope to a specific person (whom they did not know). After the envelope was delivered, the perpetrator received a parcel containing a cosmetic of small value and signed a receipt form with a fictitious name. They were to try to disguise their handwriting.

All subjects were then given a polygraph examination by an examiner blind to their guilt or innocence. A four-channel Lafayette polygraph and Reid Control Question procedures were used. Respiration, cardiovascular activity and electrodermal activity were recorded. Examiners made subjective decisions as to a subjects guilt or innocence. Examiners knew that each group of four subjects was involved in the same case and tested all subjects in a group before selecting the perpetrator. The experience of the polygraph examiner is not discussed.

Independent of the polygraph test, three other types of evidence were collected including handwriting analysis of the receipt by a handwriting expert, fingerprint analysis (from fingerprints obtained from the envelope and instruction sheet) by a fingerprint expert, and eyewitness identification by the person receiving the envelope (those people had prior knowledge that they would be required to make the identification). The handwriting expert, the eyewitness, and the fingerprint expert were to select the perpetrator from each group of four.

Results will be presented by cases instead of individuals since the data from subjects involved in each case was not independent.

	<u>Correct</u>	<u>Error</u>	<u>Inconclusive</u>	<u>Accuracy</u>
Polygraph	18	1	1	95%
Handwriting	17	1	2	94%
Eyewitness	7	4	9	64%
Fingerprints	4	0	16	100%

Fingerprinting was the most accurate method, but yielded many inconclusive results. Polygraph and handwriting had high accuracy and few inconclusives. Eyewitness identification was neither accurate nor useful. It should be noted that the fingerprints obtained in many cases were not of high quality. The authors note that the forced choice procedure used in the study would be expected to increase the detection rates and caution against overgeneralization.

Laboratory Abstracts
Relevant-Irrelevant Techniques

Richard H. Blum and William Osterloh. "The Polygraph Examination as a Means for Detecting Truth and Falsehood in Stories Presented by Police Informants." Journal of Criminal Law, Criminology and Police Science 59 (1968): 133-137.

Procedure

A total of 17 male and 3 female informants who had given information on criminal cases to local or federal agencies were selected and paid to engage in the experiment. Great care was taken to protect their identities and the fact they were cooperating with the police, as their lives would otherwise be in danger. Some of the stories that the informants told the examiner were completely true (n=9), and some were completely false (n=4). Some were partly true and partly false (n=7). (This is typical of informants, and similar to the problem in preemployment screening.) The topics about which the informants gave information covered a wide variety of crimes. The true stores were those which had been provided to their police handler on a previous occasion which had stood the scrutiny of investigation. A false story was one jointly invented by the officer and the informant, but compatible with the informant's ordinary role and opportunities, and containing credible information. True stores with false items were true stores told in the past (which had been confirmed by investigation) in which some critical items had been made false such as the name of the offender, the place where the goods were hidden or fenced, etc. Each story was first written, then rehearsed with the police handler. The examiner had no idea as to which examinee was telling the truth, which was lying all together, and which was only partly truthful. He also was blind as to the ratio of truth telling and lying. Seventeen stories contained 5 relevant questions; the remaining 3 stories contained 6, 7 and 8 relevant questions.

Three professional polygraph examiners used a standard three-channel Stoelting polygraph and the relevant-irrelevant technique. They were forced to make a decision on all of the 106 relevant questions from 20 subjects. Inconclusive decisions and reexaminations were not permitted as might occur in regular practice. Subjects were instructed not to make admissions during the polygraph examination. The physiological measures recorded are not discussed. Examiners made a subjective judgment of truth or deception on each critical item of each story.

Results

Examiners were able to identify stories as totally true, totally false, or stories containing some false information (irrespective of the particular question) as illustrated in the following table:

		<u>ACTUAL</u>		
		<u>Totally False</u>	<u>Totally True</u>	<u>Partially False</u>
<u>Predicted</u>	<u>Totally False</u>	4	0	0
	<u>Totally True</u>	0	9	0
	<u>Partially False</u>	0	0	7

However, the examiners were not always able to correctly classify individual relevant questions of informants telling partially true stories. They were incorrect about 4 questions involving 3 subjects. Thus, 102 of 106 questions were correctly classified.

Eileen J. Correa and Henry E. Adams. "The Validity of the Preemployment Polygraph Examination and the Effects of Motivation." Polygraph 10 (3) (September 1981): 143-155. (Also-Eileen Joyce Israel. Unpublished Doctoral Dissertation, University of Georgia, Athens, Georgia, 1977.)

Procedure

Twenty male and 20 female undergraduates at the University of Georgia participated in a laboratory study to determine the validity of a preemployment polygraph test. An equal number of males and females were randomly assigned to four groups: (1) lying with motivation, (2) lying without motivation, (3) truthful with motivation, and (4) truthful without motivation.

Each subject received two sessions. During the initial screening session, subjects filled out a Preemployment Data Sheet with an experimenter; nine potential lie questions were identified. Three of the nine potential lie questions (control lies) were included in case subjects had initially lied while filling out the Preemployment Data Sheet. These control lies involved activities occurring during the initial screening session.

Subjects in the lying groups were to lie to the nine questions; subjects in the truthful groups were to answer all questions truthfully. Subjects in the motivated groups were told that the best liar in the lying group and the best truth teller in the truthful group would get a \$25 reward and that the award selection would be based on the polygraph charts.

The polygraph test used a relevant/irrelevant technique. The test consisted of three series of questions about the Preemployment Data Sheet. Each series contained three potential lie questions (one control lie in each series). Each question from the Preemployment Data Sheet either preceded or followed an irrelevant norm question. Between charts the examiner questioned subjects about any responses that appeared to be larger to the relevant questions than the irrelevant questions. These questions were rephrased and inserted at the beginning of the next series. The examiner (experience not discussed) made a subjective decision of lying or truthful for each subject and for each question. The examiner was blind as to subject status. Respiration (thermister inside nostril), heart rate, and skin resistance were recorded.

Results

The following table presents the data on the examiner's overall judgement of the subject:

	<u>Actual</u>			
	<u>Lying</u>		<u>Truthful</u>	
	<u>Motivated</u>	<u>Unmotivated</u>	<u>Motivated</u>	<u>Unmotivated</u>
<u>Predicted</u>				
<u>Lying</u>	10	10	0	0
<u>Truthful</u>	0	0	10	10

The examiner correctly classified all of the subjects.

The individual lie questions also were analyzed. Three of the nine lies were control lies; the remaining six lies were from the Preemployment Data Sheet. All of the control lies were correctly identified; there were some errors in the identification of the other lies as indicated in the following table:

		<u>Correct</u>	<u>Incorrect</u>
<u>Control Lie Questions</u>	<u>Lying-Motivated</u>	30	0
	<u>Lying-Unmotivated</u>	30	0
<u>Data Sheet Lies</u>	<u>Lying-Motivated</u>	49 (82%)	11 (18%)
	<u>Lying-Unmotivated</u>	41 (68%)	19 (32%)

The difference between the motivated and unmotivated groups was not significant.

Objective numerical analyses of the data also were computed. There was no overall motivation effect, but there were some interactions between motivation and other independent variables. Lying responses were accompanied by increases in skin conductance level and large decreases in heart rate, whereas truthful responses were accompanied by decreases in skin conductance level and small decreases or increases in heart rate.

Laboratory Abstracts

**Guilty Knowledge
and
Peak of Tension
Techniques**

Kristen D. Balloun and David S. Holmes. "Effects of Repeated Examinations on the Ability to Detect Guilt with a Polygraphic Examination: A Laboratory Experience with a Real Crime." Journal of Applied Psychology (1979) 64(3): 316-322.

Procedure

Eighteen male college students with high Pd (psychopathic deviate scores on the Minnesota Multiphasic Personality Inventory) and 16 subjects with low Pd scores participated in a study to determine the effect of repeated examinations on detection of deception. Each subject was tested with two confederates. Subjects were told that the purpose of the experiment was to determine whether they did better on a bogus intelligence-information test when working alone or in groups, and that they would work in groups. Subjects were told it was important to do well and if their performance was poor they would have to meet with a panel of psychologists. When the experimenter left the two confederates urged the subject to cheat. One of the confederates said a friend had previously taken the test and produced a copy of the answer sheet. About 50% of the subjects actually cheated.

Subjects were taken to the laboratory and told that it had been learned that some had cheated. They were administered the guilty knowledge test after the test procedures were explained. The test consisted of two series of questions, each series containing 5 relevant questions, and each of those 5 relevants had 5 alternatives. The relevant questions contained information that only guilty subjects could have known from the answer sheet provided by the confederate. The first alternative to each question was a buffer. Heart rate, skin resistance, and finger pulse volume were recorded. If the critical alternative to each question elicited the largest response it was scored 2; if the critical item elicited the second largest response it was scored 1; all other possibilities were scored 0. Scores were summed across questions. Subjects with scores greater than 5.5 were called cheaters; those with scores less than 5.5 were called noncheaters.

Results

Results for skin resistance are in the following table:

		<u>Actual</u>			
		<u>Series 1</u>		<u>Series 2</u>	
		<u>Cheaters</u>	<u>Noncheaters</u>	<u>Cheaters</u>	<u>Noncheaters</u>
<u>Predicted</u>	<u>Cheaters</u>	11	2	3	1
	<u>Noncheaters</u>	7	14	15	15

On series 1, 61% of the cheaters and 87.5% of the noncheaters were correctly identified. On series 2, 17% of the cheaters and 94% of the noncheaters were correctly identified. There was a significant reduction in the detection of cheaters from series 1 to series 2.

Analysis of variance on the scores for skin resistance revealed significantly higher scores on test 1 than test 2 and significantly higher scores for cheaters than noncheaters. There were no significant effects or interactions for Pd (psychopathic deviance). Heart rate and finger pulse volume were not effective in detecting cheating.

P.O. Davidson. "Validity of the Guilty - Knowledge Technique: The Effects of Motivation," Journal of Applied Psychology 52(1) (1968): 62-65.

Procedure

Forty-eight college students (sex not specified) were divided into 12 groups to determine the effects of motivation on the detection of deception. In each group there were 3 thieves and 1 innocent subject. Each thief was to attempt to steal a sealed envelope containing a voucher with a value up to \$50.00 from a victim. The innocent subject had no knowledge that a crime was to be committed. There were 12 victims and 12 crimes, one for each group.

The study was designed to maximize the involvement of the participants. The thieves worked individually and personally planned the details of their crime. They were told that there were others after the same victim and the first one successful would get the envelope. The voucher could be cashed if the thief deceived the polygraph examiner. There also were to be prizes for the most ingeniously devised crimes whether successful or not. Motivation was varied by size of the voucher. Half of the victims had vouchers ranging from \$25.00 to \$50.00 while the other half had vouchers worth \$0.10 to \$1.00. The thieves did not know the value of the voucher held by their victim unless the envelope was obtained. Each thief in each group thought obtaining the envelope was possible. In reality, the situation for each group was structured so that one thief tried but failed, one had no opportunity to attempt the crime, and one who committed the crime. Only the thieves that completed the crime were considered guilty. Thus, there were 12 guilty and 36 innocent subjects.

Subjects were administered a polygraph test using the guilty-knowledge technique. The experience of the polygraph examiner was not discussed. Each test consisted of 6 multiple-choice questions each with 5 alternatives about the crime. Subjects were to listen to all questions and make no verbal response. The first alternative on each question was not scored. The victims had supplied the examiners with some of the details of the crimes, but the examiner did not know the correct answer to each question or who was the criminal. The criminal was the only one who could recognize the correct answer to each question.

Electrodermal activity, respiration, and heart rate were recorded, but only electrodermal activity was scored. Scoring was done without knowledge of the correct answer. Each subject's electrodermal responses to each question were ranked in order of amplitude. If the largest response was to a relevant alternative the question was scored 2. If the second largest response occurred to the relevant question it was scored 1. Other possibilities were scored 0. Subjects with a score of 6 or less were considered innocent. Subjects with a score greater than 6 were considered guilty.

The following results were obtained:

		<u>ACTUAL</u>	
		<u>Guilty</u>	<u>Innocent</u>
<u>Predicted</u>	<u>Guilty (high value)</u>	6	0
	<u>Guilty (low value)</u>	5	0
	<u>Innocent</u>	1	36

The difference between motivation groups was not significant. The authors noted that their study used a very involved mock crime with some of the uncertainties of real crimes.

Miroslav Dufek, Department of Criminalistics Legal Faculty, Charles University, Prague. "A Contribution on the Problem of Polygraph Examinations." Czechoslovak Criminalistics (February 1969). (Tr. from Czech.)

Procedure

Several experiments were conducted investigating detection of deception with guilty information type tests. The authors state they worked with equipment from the Alvar firm and recorded electroencephalograms, electromyograms, electrodermal activity, etc.

Procedure 1

In the first experiment 30 individuals were told to write a random number from 1 to 10 on a piece of paper which they retained during the test. Chance was 10 percent. The experimenter utilized a searching peak of tension test in which the opening words were always (Czech, the equivalent of) "Did you write the number...?" The answer was "No."

Results

Using a searching peak of tension test, the examiner was able to pick the correct number with certainty in 25 of the examinations (83 percent). In three of the remaining five cases, the examiner could not distinguish between two numbers, one of which was correct. In two of the examinations, the number could not be determined.

Procedure 2

The second experiment involved 20 men who examined in detail one of six objects placed before them, then concealed from the examiner the name of the object they handled (a lighter, cigarette, pencil, pen, eraser, or ashtray). The object was to be selected at random from the group. The examiner was blind to the item selected. Chance was 17 percent.

Results

Using a searching peak of tension test, in which the subject did not know the sequence of questions, the examiner detected the correct object in 18 of 20 examinations (90 percent). There were no reactions on which to make a judgment for two other subjects, an inconclusive rate of 10 percent.

Procedure 3

Five men and five women were told to select someone with whom they were closely associated with their personal life, an association of an intimate nature. With the help of an assistant, another 10 first names were selected which had no emotional meaning to the subject. The examiner was blind as to the meaningful name. During the searching peak examination, the subjects denied a relationship with all of the names. Chance was 9 percent.

Validity

The examiner determined the correct name in all ten cases (100 percent).

Procedure 4

Experiment 4 used the same procedures as Experiment 2. The difference was that in this series of peak of tension examinations, the test individual knew in advance the exact order in which the objects would be presented. The purpose was to develop progressive stress, as the correct object was approached, and relaxation following the correct item. Twenty men were tested.

Validity

In these experiments the correct item was selected in 17 of the 20 examinations (85 percent), while three of the examinations did not yield clear reactions to a single item. Instead, in these three cases, there were responses to two of the items, of which one of them was correct.

Jana Kronbergerova and Miroslav Dufek. "Experimental Experiences With the Use of Polygraph." Socialist Legality (May-June 1969). (Tr. from Czech.)

Procedure

Ten men (aged 21 to 47 years) were instructed to hide a weapon previously used in a crime of violence in any one of 18 locations. In a subsequent polygraph examination, questions were structured so that it would be possible to get a general specification (such as the building) as well as the floor number and the room of the hidden weapon. The instrument recorded respiration, electrodermal responses, and blood pressure. The first and last questions were control questions (i.e. places that were not possible locations); the remaining questions involved possible hiding places. The number of questions asked was not stated, and it was reported that none of the questions were repeated. Subjects were required to deny the real location. Chance was six percent for detecting the complete location. The method of scoring the polygraph tests is not reported. Examiners are said to be experienced and educated in psychology, physiology and other fields. The authors give insufficient detail about how the polygraph test was conducted.

The experimenters were able to determine the exact location of the weapon in nine of the ten cases. In the tenth, they were able to determine the building and floor, but not the correct room. The examinee reacted more strongly to another room than he did to the room in which he hid the weapon. In retrospect it appeared that he had a strong personal reason to react to the other room.

David T. Lykken. "The GSR in the Detection of Guilt." Journal of Applied Psychology 43 (6) (1959): 385-388.

Procedure

Forty-nine male college students were assigned at random to four groups. Those in Group 1 (n=13) enacted two mock crimes in random sequence, a "murder" and a "theft." Both crimes required some rather detailed play-acting. Those in Group 2 (n=12) enacted only the murder, and those in Group 3 (n=12) only the theft. Those in Group 4 (n=12) were not exposed to either crime. Each subject was told that he was to be questioned in relation to two crimes. The subject was told that if the experimenter (experience is not discussed) felt that the subject's electrodermal response indicated guilt to any question, the subject would receive an electric shock. An unpleasant shock was then demonstrated. In fact, shocks were given during testing of all the subjects to heighten general anxiety. Each test consisted of two lists of six multiple choice questions - one for the murder and one for the theft. The experimenter read the question and then read each of several alternative answers, with adequate time between them for GSR recovery. The six questions related to details of the crimes known only to the perpetrators (guilty knowledge). The subject was not to reply to any questions. The position of the critical item in each list was varied from one subject to another, and the examiner was in the blind as to which item was critical in each of the lists and was blind to the subject's group. Electrodermal responses to the alternatives in each question were ranked based on amplitude. If the largest response occurred to the relevant alternative it was scored 2; if the second largest response occurred to the relevant alternative a score of 1 was given; otherwise, a score of 0 was given. Scores were summed across questions. A subject with a score of 6 or less was classified innocent; a subject with scores over 6 was classified guilty.

Results

Four subjects from Group 1 and one subject from Group 2 were misclassified as to group. There were six errors in the 98 interrogations (each subject had two interrogations, one for each crime); all of these errors involved classifying a subject innocent when he was really guilty. The following table was constructed from Lykken's presented data:

		<u>Actual</u>		
		<u>Guilty of Two Crimes</u>	<u>Guilty of One Crime</u>	<u>Innocent</u>
<u>Predicted</u>	<u>Guilty of Two Crimes</u>	9	0	0
	<u>Guilty of One Crime</u>	3	23	0
	<u>Innocent of Both Crimes</u>	1	1	12

David T. Lykken. "The Validity of the Guilty Knowledge Technique: The Effects of Faking." Journal of Applied Psychology 44(4) (1960): 258-262.

Procedure

Twenty subjects (males and females) participated in a study on the validity of the guilty knowledge technique. Each subject initially completed a 25-item questionnaire requiring biographical information. Subjects were subsequently given a test of guilty knowledge using the 25 questions each with 6 alternatives. One alternative pertained to the subject; 4 alternatives to each question pertained to other subjects. The first alternative was a randomly selected item and was not scored to reduce the effects of orienting. The examiner (experience not discussed) attempted to identify the subject's set of questionnaire responses.

Electrodermal activity was measured. Each subject was given a 15-minute lecture on electrodermal activity, lie detection in general, and the principle of the guilty knowledge method in particular. Each subject was allowed to practice (with feedback) producing voluntary electrodermal responses by various methods. The group was told of test format, cautioned not to attempt to defeat the test by inhibiting responses, and advised that the best way to defeat the test would be to produce electrodermal responses of various amplitudes to the innocent alternatives as in random a pattern as possible. Each subject was offered a \$10.00 prize if they could defeat the test.

Electrodermal responses to the 5 alternatives (omitting the first item) were ranked based on amplitude. Thus, there was a distribution of ranks associated with each of the 5 questionnaires used for each subject. Each distribution was then given a score by giving a point every time a particular rank occurred more than 5 times. A distribution in which all 25 responses had the same rank would be scored 20. A distribution with each rank occurring 5 times would be scored 0. When no measurable electrodermal response occurred an item was not ranked. The scores for each distribution were compared. Patterns of nonrandom responding (higher scores) were considered indicative of guilty knowledge.

Results

The questionnaire belonging to each subject was correctly identified in all 20 cases. The author notes that this study uses a more comprehensive interrogation than his previous study (Lykken, 1959).

Kazuo Ohnishi, Katsunori Matsuno, Masana Arasuna and Akhiro Suzuki. "The Objective Analysis of Physiological Indices in the Field of Detection of Deception." Reports of the National Institute of Police Science 29(3) (August 1976): 181-188. Abstract in English.

Procedure

A searching peak of tension test to detect the correct number from six numbers (unknown to the examiner) was conducted during the polygraph examination of 50 suspects in criminal cases in the laboratory of the Osaka Prefecture Police Headquarters. The charts were analyzed numerically at the National Institute of Police Science applying a criterion of rank scores. Chance on each test was 17 percent. The location of the number in the list was unknown to the blind reviewer and no dummy number was inserted at the beginning of each list.

Results

Applying only the rank score of breathing suppression, the detection rate was 46 percent. Applying only the rank score of electrodermal responses, the correct detection of the critical number was 72 percent. Detection with each individual measure was significantly greater than chance. When the two indices were combined, the detection rate was 92 percent.

Robert M. Stern, John P. Breen, Takami Watanabe, and Bradley S. Perry. "Effect of Feedback of Physiological Information on Responses to Innocent Associations and Guilty Knowledge." Journal of Applied Psychology, 66 (6) (1981): 677-681.

Procedure

Experiment 1 involved 24 males and 24 females to determine the effect of feedback on innocent associations. Subjects selected a geometric figure from five possible figures and wrote down their social security number and four additional social security numbers from a provided list. Sixteen subjects received no feedback, 16 received auditory heart rate feedback, and 16 auditory electrodermal feedback while being asked four series of questions. In two series, subjects were asked if they had picked each of five geometric forms. In another two series, subjects were asked if each social security number was theirs. Subjects were required to respond with no to all questions. The first question of a series was a buffer question to reduce orienting. Detection was based on the electrodermal data with the response to each of the five questions in each series scored from one to five based on magnitude. The mean rank to each question of the two social security number series and the mean rank to each question of the two geometric form series was then computed.

The proportion of subjects showing the largest electrodermal response to the relevant stimulus (their social security number and the selected geometric form) significantly varied with feedback and type of stimulus and is as follows:

(1) No feedback - form (5/16), number (9/16); (2) Heart rate feedback - form (8/16), number (11/16); (3) Electrodermal feedback - form (11/16), number (12/16). Thus, subjects receiving electrodermal feedback were easiest to detect and social security numbers were easier to detect than geometric forms.

In Experiment 2, 26 males and 26 females were subjects in a mock crime situation. All subjects played the role of an American Embassy hostage in Iran. The 26 designated guilty subjects received written instructions (which they were to read three times) about their involvement in a plot to murder the Ayatollah. The 26 innocent subjects received instructions that had some of the same relevant words as those of guilty subjects but involved no murder plot. Thirteen innocent subjects and 13 guilty subjects received auditory electrodermal feedback. The remaining subjects received no feedback. The guilty knowledge technique was used. Ten questions, each with five alternatives, were presented to the subject with the first alternative serving as a buffer and not scored. Four of the ten questions were control questions. Subjects were to make no response to the questions. Responses were scored by the method of Lykken (1959).

Results

Results are demonstrated in the following table:

	<u>Guilty</u>	<u>Actual</u>	<u>Innocent</u>
<u>Guilty (No Feedback)</u>	12		2
<u>Guilty (Feedback)</u>	13		1
<u>Predicted</u>			
<u>Innocent (No Feedback)</u>	1		11
<u>Innocent (Feedback)</u>	0		12

Thus, 96% of the guilty and 88.5% of the innocents were correctly classified. Electrodermal feedback did not significantly affect detection accuracy.

William M. Waid, Emily Carota Orne, Mary R. Cook, and Martin T. Orne. "Effects of Attention as Indexed by Subsequent Memory On Electrodermal Detection of Information." Journal of Applied Psychology. (1978) 63(6): 728-733.

Experiment 1

Procedure

Twenty-nine designated guilty subjects and 11 designated innocent subjects (males) participated in a study to determine the effect of attention on detection of deception using a guilty knowledge test. Guilty subjects learned six code words during a series of tasks. Innocent subjects completed a questionnaire during this period. Guilty subjects were told that highly intelligent and mature individuals could deceive the examiner. Care was taken to involve innocent subjects in the procedure. All subjects were told that the way to appear innocent was not to respond during the test. Subjects were tested by an experimenter blind to subject status. Subjects were presented twice a list of 24 words. There were 6 categories of words with each category containing one of the learned code words. A word was considered detected if the code word elicited a larger response than the remaining 3 words in the category. For innocent subjects, one of the four words in each category was arbitrarily designated as the critical word. One category of four words was eliminated from the analysis. Skin resistance was measured. After the polygraph test, subjects were given a list of 115 words and asked to indicate the words heard during the polygraph test.

Results

When the criterion for detection was 3 or more words (chance = 2.5 words), the following results were obtained:

		<u>Actual</u>	
		<u>Guilty</u>	<u>Innocent</u>
<u>Predicted</u>	<u>Guilty</u>	23	3
	<u>Innocent</u>	6	8

Seventy-nine percent of the guilty subjects and 73% of the innocent subjects were detected. Guilty subjects had significantly more critical words detected than innocent subjects.

Guilty and innocent subjects did not differ in percentage of words recalled. Detection in innocent and guilty groups did not appear related to number of words recalled although there were some trends.

Experiment 2

Procedure

Procedure was the same as Experiment 1 with some minor differences. Eighteen

guilty and 10 innocent subjects (males) were used. The list of 24 words was presented 5 times. Subjects were tested in groups of 3 to 7 by an examiner in another room. Subjects received the stimuli through headphones and indicated their answers by moving their head. The answer was observed through a one-way window. Subjects were blindfolded so they could not see the other subjects and were told they would receive a shock at the end of the test proportional in strength to the number of detected lies. After the test, subjects were to write down as many of the words involved in the test they could recall. A code word was considered detected if it elicited a larger response than the other three codewords in its category. One category of 4 words was removed from all analysis.

Results

Subjects were called guilty if seven or more critical words were detected (chance = 6 words). The following results were obtained:

		<u>Actual</u>	
		<u>Guilty</u>	<u>Innocent</u>
<u>Predicted</u>	<u>Guilty</u>	11	1
	<u>Innocent</u>	7	9

Sixty-one per cent of the guilty subjects and 90 percent of the innocent subjects were correctly detected.

Guilty subjects had significantly more critical words detected than innocent subjects and guilty subjects recalled significantly more words than innocent subjects. Among guilty subjects correlations between number of words recalled and number of words detected was .48. Detected guilty subjects recalled significantly more words than not detected guilty subjects.

Experiment 3

Procedure

Procedure was the same as Experiment 1 except for minor differences. Fifteen innocent and 15 guilty subjects (males) were tested. The list contained 36 words with 6 words in each category; 1 word in each category was the code word. The list was presented 4 times. One category of 6 words was eliminated from all analyses. Skin conductance was measured. As in Experiment 2, subjects were to write down as many words they could recall after the test.

Results

The following results were obtained when the criterion for detection was 6 words (chance = 5 words):

		<u>Actual</u>	
		<u>Guilty</u>	<u>Innocent</u>
<u>Predicted</u>	<u>Guilty</u>	11	3
	<u>Innocent</u>	4	12

Seventy-three percent of the guilty subjects and 80% of the innocent subjects were detected. Guilty subjects had significantly more code words detected than innocent subjects but there was no difference between guilty and innocents in number of words recalled. Among guilty subjects, number of words recalled was correlated .41 with number of words detected. Detected guilty subjects recalled significantly more words than not detected guilty subjects.

William M. Waid, Emily C. Orne, Mary R. Cook, and Martin T. Orne. "Meprobamate Reduces Accuracy of Physiological Detection of Deception." Science, 212 (1981): 71-73.

Procedure

Male college students (aged 18 to 24 years) were paid \$2 an hour as subjects in an experiment to determine the effect of meprobamate on detection of deception. Subjects designated guilty overlearned six words during timed, interpreted tasks prior to the polygraph test. Innocent subjects performed the same tasks but learned no words. Guilty subjects were told that it was possible to deceive the examiner, whereas innocent subjects were told it is often difficult to prove one's innocence. Guilty subjects were randomly assigned to one of three groups (11 per group). One group was told they were taking a tranquilizer which would help them avoid detection and were given 400 mg. meprobamate. Another group was given a placebo. The third guilty group received nothing.

After 30 minutes, a second experimenter, blind to subject status, administered a polygraph test using the guilty knowledge technique. The question list consisted of 24 words, four in each of six categories. One of four words in each category was a word the guilty subject memorized. The first word of the list was a dummy word. The list was presented twice. The examiner also completed a rating scale about his perception of subject's drug status.

Electrodermal response amplitude, smallest inspiration, and change in relative blood pressure were recorded and scored by technicians who had not observed the test. A subject was classified guilty if his response to the relevant word in each set of four produced a larger response than the remaining three words on five or more sets. In order to analyze data from innocent subjects, one word in each set of four was randomly designated as the critical word.

The following results for each measure were obtained:

		<u>Actual</u>			
		<u>Guilty: No Pill</u>	<u>Innocent</u>	<u>Guilty: Placebo</u>	<u>Guilty: Meprobamate</u>
<u>Electrodermal</u>					
	Guilty	9	0	8	3
	Innocent	2	11	3	8
<u>Cardiovascular</u>					
<u>Predicted</u>	Guilty	0	1	2	0
	Innocent	11	10	9	11
<u>Respiratory</u>					
	Guilty	4	1	3	5
	Innocent	7	10	8	6

Correct classification was reduced significantly in the meprobamate group. Electrodermal activity was a better discriminator between guilty and innocent than the other measures.

The authors note that results were not due to reduced electrodermal responsiveness in the meprobamate group since there were no differences between the groups in average number of critical words eliciting a measurable response and no differences in average response to all questions. Examiners were not able to accurately judge whether or not subjects had received meprobamate.

Kazunoba Yamaoka and Akihiro Suzuki. Studies on Skin-Blood Flow as an Index of Lie Detection. Polygraph 9(4) (December 1980): 232-237*.

Procedure

Seven females and 6 males (aged 18 to 29 years) were subjects in a study to determine the effectiveness of skin-blood flow as an index of lie detection. Subjects selected 1 of 5 playing cards. When the subject was asked about each card during a subsequent polygraph test he was to reply no to all questions. The examiner was blind to the selected card; experience of the examiner is not discussed. Subjects were presented 2 series of questions about each card; each series contained a different question order. The response to the first question was eliminated. Skin potential, skin resistance, and skin-blood flow were recorded. Responses to each question for each system were ranked from 1 to 5 with the largest response receiving a rank of 1. The following results were obtained:

	<u>Number of Successful Detections</u>	<u>Number of Unsuccessful Detections</u>
<u>Skin-Blood Flow</u>	5	8
<u>Skin Resistance</u>	2	11
<u>Skin Potential</u>	10	3

Skin potential was the only response measure able to detect the selected card significantly greater than chance levels.

*Previously published in Reports of the National Institute of Police Science
26 (1973): 206-209. Text in Japanese.

References Cited

References Cited

Most of the references cited in the text will be found in the abstracts section. Those references which are not in the abstracts are listed below:

Ash, Philip. "Survey of Attitudes on the Polygraph." Polygraph 2(3) (September 1973): 200-223.

Barland, Gordon H. "On the Accuracy of the Polygraph; an Evaluative Review of Lykken's Tremor in the Blood." Polygraph 11(3) (September 1982): 258-272.

Beijk, J. "Experimental and Procedural Influences on Differential Electrodermal Activity." Psychophysiology 17(3) (1980): 274-278.

Carlucci, Frank. Memorandum for Secretaries of Military Departments, Chairman, Joint Chiefs of Staff, Directors of Defense Agencies. Subject: "Degradation of Operational Readiness/Mission accomplishment Due to Personnel Security Investigative Shortfalls." 6 August 1982.

"A Comparison of Information Obtained during Military Interview Program Interviews with that Obtained during Military Conversion Polygraph Interviews." unpublished manuscript, National Security Agency, 1983.

Dawson, Michael E. "The Delayed Answer Test (DAT) and the Effects of Countermeasures." Paper presented at the Annual Meeting of the American Polygraph Association, Las Vegas, August, 1977.

Dawson, Michael E. "Physiological Detection of Deception; Measurement of Responses to Questions and Answers During Countermeasure Maneuvers." Psychophysiology 17(1) (January 1980): 8-17.

Engel, B. T. "Stimulus-Response and Individual-Response Specificity." Archives of General Psychiatry 2(1960): 305-313.

Garwood, M.; Engel, B. T.; and R. Capriotti. "Autonomic Nervous System Functioning and Aging: Response Specificity." Psychophysiology 19(4)(1982): 378-385.

Gustafson, Lawrence A. and Orne, Martin T. "Effects of Heightened Motivation on the Detection of Deception." Journal of Applied Psychology 47(1963): 408-411.

Gustafson, Lawrence A. and Orne, Martin T. "The Effect of 'Lying 'in 'Lie Detection' Studies." Paper read at the 35th Annual Meeting of the Eastern Psychological Association, April 17, 1964, Philadelphia, Pennsylvania.

Gustafson, Lawrence A. and Orne, Martin T. "The Effects of Verbal Responses on the Laboratory Detection of Deception." Psychophysiology 2(1)(1965): 10-13.

Heckel, R. V.; Brokaw, J. R.; Sulzberg, H. C. and Wiggins, S. L. "Polygraph Variations in Reactivity Between Delusional, Non-Delusional, and Control Groups in a 'Crime' Situation." Journal of Criminal Law, Criminology and Police Science 53(3)(1962): 380-383.

Kugelmass, Sol and Lieblich, Israel. "Relation Between Ethnic Origin and GSR Reactivity in Psychophysiological Detection." Journal of Applied Psychology 52(2)(1968): 158-162.

Kugelmass, Sol; Lieblich, Israel; Ben-Ishai, Akiva; Opatowski, Abraham and Kaplan, Maier. "Experimental Evaluation of Galvanic Skin Response and Blood Pressure Change Indices During Criminal Interrogation." Journal of Criminal Law, Criminology and Police Science.

Lieblich, Israel. "Manipulation of Contrast Between Differential GSR Responses Through the use of Ordered Tasks of Information Detection." Psychophysiology 6(1)(1969): 70-77.

Lieblich, Israel; Kugelmass, Sol; and Ben-Shakar, Gershon. "Efficiency of GSR Detection of Information as a Function of Stimulus Set Size." Psychophysiology 6(5)(1970): 601-608.

Lieblich, Israel; Naftali, Gideon; Shmueli, Joseph and Kugelmass, Sol. "Efficiency of GSR Detection of Information with Repeated Presentation of Series of Stimuli in Two Motivational States." Journal of Applied Psychology 59(1)(1974): 113-115.

Lindsey, Robert. The Falcon and the Snowman: A True Story of Friendship and Espionage. New York: Simon & Schuster, 1979.

Linehan, John G. "An Aspect of World War II Use of the Polygraph." Polygraph 7(3)(September 1978): 233-239.

Lykken, David T. "The Detection of Deception." Psychological Bulletin 86(1)(1979): 47-53.

Lykken, David T. "The Psychopath and the Lie Detector." Psychophysiology 15(2)(1978): 137-142.

Lykken, David T. A Tremor in the Blood, Uses and Abuses of the Lie Detector. New York: McGraw-Hill, 1980.

MacNitt, Reginald D. "In Defense of the Electrodermal Response and Cardiac Amplitude as Measures of Deception." Journal of Criminal Law, Criminology and Police Science 33(1942): 266-275.

Marston, William Moulton. The Lie Detector Test. New York: R. R. Smith, 1938.

Miyake, Yoichi. "A Study of Skin Resistance Response, Photoplethysmographic Vasomotor Response and Eye Movement as Indices of Lie Detection." Reports of the National Institute of Police Science 31(2)(May 1978): 18-24.

Ohkawa, Hisatsugi. "Comparison of Physiological Response of 'yes', 'no', and 'mute' Conditions in Peak of Tension Test." Research Materials No. 21, Polygraph Reports, National Institute of Police Science (1963): 1-4.

Phannenstill, Richard J. "The Polygraph Passes the Test." Security Management 27(8)(August 1983): 58-60.

Podlesny, John A. and Raskin, David C. "Effectiveness of Techniques and Physiological Measures in the Detection of Deception." Psychophysiology 15(4)(July 1978): 344-359.

Putnam, Richard L. "Polygraph Screening of Police Applicants: Necessity or Abuse?" Polygraph 7(4)(December, 1978): 257-262.

"Security Interviews with the Aid of a Polygraph Compared to the Special Background Investigation as Sources of Information." Unpublished manuscript, National Security Agency, 1983.

Silverberg, Ben A. "Attitudes of Job Applicants Toward the Arther Security Clearance Examination." Journal of Polygraph Science 14(6)(May-June, 1980): 1-4.

Silverberg, Ben A. "Attitudes of Job Applicants and Employees Toward the Polygraph." Polygraph 9(3)(September, 1980): 162-169.

Silverberg, Ben A. "Employee Attitudes Toward Periodic Polygraph Examination." Journal of Polygraph Science 15(4)(January-February, 1981): 1-4.

Statistical Study on the Use of the Polygraph as a Personnel Security Screening Aid." Unpublished manuscript, National Security Agency, 1953.

Thatcher, Margaret. "Text of Prime Minister's Statement: Security Commission Report on the Prime Case." 1982.

Trovillo, Paul V. "Report on a Survey of Methods of Operation and Accomplishments of Russell Chatham, Inc. Polygraph Program at Oak Ridge, Tennessee." Unpublished manuscript, April 14, 1951. 43 pp.