

---

Chapter 3

# **Controversy Over Polygraph Testing Validity**

# Controversy Over Polygraph Testing Validity

---

## INTRODUCTION

The validity of polygraph examinations to detect deception has long been a controversial issue (cf. 108,136,194,195). Since development of polygraph techniques almost 80 years ago, their use both within and outside the Federal Government has been the focus of numerous judicial opinions and, as well, legislative and executive branch debate. Polygraph examinations have been advocated as a way to ascertain guilt of criminal suspects, to exculpate innocent suspects, to protect national security, and to maintain employee honesty. Polygraph examinations have, at the same time, been criticized for providing inaccurate and misleading information, for failing to detect security risks (167), for interfering with the rights of private citizens (128), and for lowering employees' morale. At the center of controversy over the use of polygraph examinations is the question of its validity: does a polygraph examination actually identify truthful and nontruthful individuals?

Recent interest in polygraph examinations and their validity stems from efforts to broaden Federal Government use. The Department of Defense (DOD), in late 1982, drafted revisions to existing regulations (5210.48). DOD proposed expansion of the use of polygraph tests for preemployment screening and periodic or aperiodic testing of employees who have access to highly classified information. Currently, only the National Security Agency (NSA) and the Central Intelligence Agency (CIA) are able to use polygraph tests in this way. Expanded use of polygraph testing in all Federal agencies was made explicit in a Presi-

dential National Security Decision Directive (Mar. 11, 1983, NSDD-84). In part, the directive requires agencies and departments which handle classified information to revise existing regulations to permit *use* of polygraph examinations as part of internal investigations of unauthorized disclosure of classified information. Prior to the directive, investigations of unauthorized disclosures had to be referred to the Department of Justice (DOJ). Employees who refuse to submit to a polygraph examination could, if NSDD-84 is implemented, be subject to adverse consequences. In October 1983, DOJ announced that administration policy would also permit Government-wide polygraph *use* in personnel security screening of employees (and applicants for positions) with access to highly classified information,

Proposals to expand use of polygraph examinations to maintain national security have renewed the debate about the appropriateness of various polygraph techniques and their ability to detect deception. In order to provide a context for the present evaluation of scientific evidence on the validity of polygraph testing, previous assessments of accuracy of polygraph testing are reviewed in this chapter. Legal precedents regarding polygraph testing and congressional hearings on its use, both within and outside of Government, are briefly considered. The chapter also describes scientific criteria for establishing validity and reviews other efforts to evaluate the scientific literature on testing.

## JUDICIAL REVIEWS

When courts have been called on to resolve disputes concerned with use of polygraph examinations, they have had to consider both the tech-

nique's validity and whether its use, however valid, interfaces with other values that the law seeks to protect. The varying decisions reached

by State appellate courts and Federal circuits (see 8) may in large measure reflect varying beliefs about the validity of polygraph examinations. Indeed, for many years, the leading case on the admissibility of novel scientific evidence (*Frye v. United States* (58)) was a case about the admissibility of polygraph evidence, and the opinion centered on the question of validity. The issue of how a court is to decide the question of any scientific technique's validity has brought the Frye test into question in recent years and makes salient the problem of establishing judicial standards for assessing validity (60).

### Polygraph Findings as Evidence

The *Frye* case involved a 19-year-old defendant convicted of robbery and murder. Prior to his trial, a well-known psychologist and one of the originators of polygraph testing, Dr. William Marston, administered a "systolic blood pressure test" to detect deception (e.g., 114). Dr. Marston determined, on the basis of this test, that Frye was truthful when he denied involvement in the robbery and murder. The trial judge, however, refused to permit Dr. Marston to either testify about the examination or conduct a reexamination using the blood pressure test in court,

Frye appealed his conviction on the grounds that relevant exculpatory evidence had not been admitted. The appeals court, however, concurred with the initial trial court judgment. The court reasoned that the systolic blood pressure deception test was validated only by "experimental" evidence and was not based on a "well-recognized scientific principle or discovery." The decision stated that, "while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the things from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs. Just when a scientific principle crosses the line between experimental and demonstrable is difficult to define,"

Ironically, Frye's conviction was later reversed when another man confessed to the crime, thereby providing Frye with more convincing corroboration of his denials of guilt. This did not settle the

case, however, and recent discussion of the facts of the case indicate that Frye was, indeed, guilty. The crude polygraph examination conducted by Marston, thus, appears to have yielded an inaccurate conclusion.

The *Frye* test is still used as precedent in most Federal courts. Subsequent opinions (in areas other than the polygraph) have tried to better define that line between "experimental" and "demonstrative" stages of a scientific innovation. For example, the court in *United States v. Stifel* (190) held that "neither newness nor lack of absolute certainty in a test suffices to render it inadmissible in court." In a second case, *United States v. Brown* (189), the court also seemed to be concerned with validity: "The fate of a defendant in a criminal prosecution should not hang on his ability to successfully rebut scientific evidence which bears an 'aura of special reliability and trustworthiness,' although, in reality the witness is testifying on the basis of an unproved hypothesis in an isolated experiment which has yet to gain general acceptance in its field." The *F-e* test has been held to be too high a hurdle by some trial courts, which have replaced it with the test for admissibility of expert testimony generally: "testimony by a witness as to matters which are beyond the ken of the layman will be admissible if relevant and the witness is qualified to give an opinion as to the specialized area of knowledge" (190).

A closely related question for the courts has been who should determine whether some procedure has gained general acceptance in its field. Some have held that the courts must look to the judgment of the scientific community (e.g., 191). In other decisions, the court refused to "surrender to scientists the responsibility for determining the reliability of (scientific) evidence," and that "a determination of reliability cannot rest on a process of 'counting (scientific) noses.'" "

Saks and Van Duizend (145) concluded that whichever set of tests is employed, the courts are in a weak position to assess validity directly or to count scientific noses. The result has been: 1) general deference by the courts to the judgments of scientific communities; and 2) "numerous incongruities . . . where less reliable scientific and technological information is admitted but the ad-

mission of demonstrably more reliable techniques is delayed until the requisite consensus has formed" (145; see, also, 60).

When the courts examine polygraph testing, they are faced with a series of dilemmas. To which "particular field" of expertise can the courts turn: physiology, psychology, polygraph? If they look to the data themselves, what are they to make of it? As the present report suggests, validity assessment involves a complex situation and technique-specific answer. Even if a final, single accuracy rate could be established, how should a court use it. How accurate must a diagnostic or predictive technique be to be deemed valid for evidentiary purposes? Regularly admitted psychiatric evidence is widely recognized (including by the U.S. Supreme Court, see *Addington v. Texas*, (2)) as having accuracy rates comparable to flipping coins (e.g., 55,208). In *Barefoot v. E'stelle* (13) the Supreme Court acknowledged that psychiatric predictions of dangerousness and violent behavior do not exceed an accuracy level of 33 percent (see 118). Yet, this evidence was held admissible in *Barefoot* and sufficiently valid to uphold a decision to execute a convicted person.

In summary, then, the courts have found themselves disagreeing on methods to establish validity for purposes of admissibility of evidence, where the critical focus of such judgment should rest. In addition, courts are inconsistent about what decision to make on the basis of judicial findings of fact regarding the validity of a diagnostic or predictive device.

### **Laws Regulating Polygraphs in Employment Settings**

As described in chapter 2, screening employees is the most frequent application of polygraph testing. Many employers argue that use of polygraph testing for preemployment screening, periodic checking, and to resolve actual thefts is necessary. Internal crime has been estimated to cost private industry up to \$10 billion annually (see 172), and polygraph testing is regarded as a cost-effective tool. Employers argue that screening applicants, and periodic checking of employees, are the most efficient ways to control pilferage, embezzlement, poaching, and other forms of theft. The need for

polygraph testing is felt particularly in industries which have high risk of theft and fraud (e. g., commercial banks), high turnover (supermarkets, other retail operations), or both.

According to Ansley (8), the use of private polygraph testing is limited by statute in 18 States plus the District of Columbia. Most of these laws seek to protect employees from being requested, required, demanded, or subjected to polygraph examinations by their employers. Employers are reported to be able to find ways around these laws. For example, employers may tell the employee that they suspect them of theft, but that if the employee can find a way to demonstrate innocence, the employer will not discharge the employee. In addition to polygraph validity, other polygraph-related concerns include issues of voluntariness, invasions of privacy, being compelled to inform on other employees, inhibiting union activity, and the polygraph as a cover for racism and sexism. This list does not exhaust concerns that have been expressed.

A survey of 143 private firms by Belt and Holden (25), regarding their use of polygraph testing, yielded a number of interesting findings. Twenty percent of respondents reported using polygraph examinations for preemployment screening, periodic surveys, and investigations of specific onsite crimes. It is interesting that of reasons given for using or not using polygraph tests, users ranked moral or ethical considerations last and efficiency first; nonusers, however, ranked validity and reliability second in importance, cost third, and the availability of qualified operators fourth in importance. The survey found a positive relationship between a State having a licensing requirement for polygraphers and employers' use of polygraph testing. According to Ansley (8), 25 States have licensing requirements for polygraphers; licensing is optional in one State.

Although there is testimony that use of polygraph testing reduces employee crime (172), no formal cost-benefit analyses appear to have been conducted. In addition, there is no research on the predictive validity of polygraph *results* (72,144). Although employee issues are critical to proposed Government uses of polygraph testing, few data are available on Government employees (see chs. 4 and 5).

One additional area of controversy has concerned employee rights and employer-employee relationships. The general matter of invasion of privacy is particularly pertinent in preemployment screening and periodic checking. In preemployment screening, the range of questions that may be asked has been subject to particularly heavy criticism. Questions have been reported to include items concerning union activity, sexual preference, and family problems (169); and, in addition, willingness to make a commitment to the job (144), and whether the respondent has ever been tempted to steal (71). During periodic checking, respondents are sometimes asked not only about their own possible improper behavior (e.g., underringing in supermarkets), but also about their level of job satisfaction, intention to remain with the employer, and activities of their fellow employees (204). There is some concern about whether prejudices of the polygraph examiner based on racial, ethnic, and gender stereotypes bias employees' responses (144). These assertions do not appear to have been researched. And no related claims under Title VII of the Civil Rights Acts have been upheld.

One argument against the use of polygraph examinations in the employment situation is that it

destroys the trust relationship between employers and employees, and creates employee dissatisfaction. However, the few employee surveys that have been conducted have not supported this argument. Apparently, five studies have examined whether the use of the polygraph causes private sector employees to be dissatisfied. In one study (144), 96 percent of applicants were willing to take a polygraph examination to get a job, 86 percent of the applicants thought the preemployment examination was fair, and 88 percent were willing to take it routinely as a condition of employment. A problem with the study was that applicants were surveyed immediately after taking the polygraph examination so they may have thought their responses were part of the screening process. In the one known survey of Federal employees, the Air Force (183a) surveyed individuals who had volunteered to participate in a pilot project on the use of the polygraph for counterintelligence/security examinations. About 99 percent of the respondents felt that the examination was fair, and were willing to take an examination for counterintelligence purposes.

## FEDERAL DEBATE OVER POLYGRAPH VALIDITY

Concern about and debate over Federal Government use of the polygraph have emerged at several points during the past 20 years. As shown in figure 1, the history is essentially one of legislative concern triggered by some executive branch proposal or action regarding polygraph testing. The questions raised by Congress have included constitutional and ethical as well as validity issues. However, the scientific validity and reliability of polygraph testing has been and is a central congressional concern. This chapter briefly describes the history of Federal Government involvement with the issue of polygraph validity.

### The 1960's

Congressional interest first intensified in 1963 when controversy developed over an executive

branch proposal to use lie detectors to find the source of unauthorized disclosures of sensitive or classified information, sometimes known as "leaks" (192). The then chairman of the House Committee on Government Operations asked the Foreign Operations and Government Information subcommittee to study the Federal Government's use of polygraphs. The study found that, excluding the National Security Agency and Central Intelligence Agency (for which information was classified), Federal agencies had conducted 19,796 polygraph examinations in 1963. In 1964, the subcommittee held hearings and received testimony from private polygraphers, researchers, and Federal officials. In a 1965 report (167), the House Committee on Government Operations concluded that there was no scientific evidence to support the theory of the polygraph, and that the research

Figure 1.—Debate at the Federal Level 1963-83

	1963-64	1965	1966	1967-74	1974	1975	1976	1977-78	1979	1980	1982	1983
<b>Executive branch activity</b>	Proposal to use polygraph to investigate leak	DOD Directive 5210.48 DOD research program Civil Service regulations				DOD issued revisions to Directive 5210.48				Director of Central Intelligence study investigative Scope and Adjudicative Procedures Among Intelligence Personnel Security Survey	DOD Select Panel Review of the DOD Personnel Security Program Deputy Secretary of Defense issues, memos instituting aperiodic polygraph exams for those with SCI clearance DOD prepares proposed revisions to Directive 5210.48	President issues NSDD-84 DOD prepares further draft of revised regulations DOJ announces administration policy on polygraph testing
<b>Legislative activity on Federal agency use</b>	Study and hearings by House Government Operations Subcommittee on Foreign Operations and Government Information	Report issued by House Government Operations Committee	Bills introduced which would restrict or ban Federal Government use of polygraph	Report Privacy, Polygraphs and Employment issued by Senate Judiciary Subcommittee on Constitutional Rights Study and hearings by House Government Operations Subcommittee on Foreign Operations and Government Information	Report issued by House Government Operations Committee	S. 1845 introduced; would bar Federal agencies and private employers from requesting or requiring polygraphs held on Judiciary Subcommittee on the Constitution	Hearings and report by House Permanent Select Committee on Intelligence Subcommittee on Oversight	Hearings by House Judiciary Subcommittee on Civil and Constitutional Rights on DOD proposed revisions	Hearings by House Post Office and Civil Service Subcommittee on Civil Service and House Judiciary Subcommittee on Constitutional Rights on NSDD-84 Congress passes amendment to DOD appropriations bill prohibiting changes in DOD polygraph regulations until Apr. 15, 1984 Hearings by Senate Committee on Government Affairs on NSDD-84			
<b>Legislative activity on private employer use</b>	Some bills would ban use of polygraphs by private employers	Report referred to private as well as Federal employer	Hearings held by House Education and Labor Subcommittee on Labor-Management Relations	S. 1845 introduced	Hearings held by House Education and Labor Subcommittee on Labor-Management Relations	HR2403 introduced; would ban use of polygraph by private employers						

evidence as to its accuracy was inadequate. The committee recommended that further research be conducted and training for polygraph examiners be upgraded, and that the President establish an interagency committee to study and work out solutions to problems posed by Federal Government use of polygraphs.

Later in 1965, an interagency polygraph committee of representatives from DOD, CIA, DOJ, Bureau of the Budget (now Office of Management and Budget), Office of Science and Technology (now the Office of Science and Technology Policy), and other executive agencies was established. The interagency committee concluded that: 1) there was insufficient scientific evidence concerning the validity and reliability of polygraph testing; and 2) the use of the polygraph constituted an invasion of privacy of the individual being interrogated. The committee recommended that the "use of the polygraph in the executive branch should be generally prohibited, and permitted only in special national security operations and in certain specified criminal cases" (166). The recommendations made at that time concerning personnel screening were promulgated as Civil Service regulations on regulating the use of polygraphs in personnel investigations of competitive service applicants and appointees to competitive service positions (ch. 736, app. D, of the Federal Personnel Manual). According to these regulations, which are still in effect, only executive agencies with highly sensitive intelligence or counterintelligence missions directly affecting the national security such as "a mission approaching the sensitivity of that of the Central Intelligence Agency" are permitted to use the polygraph for employment screening and personnel investigations of applicants for and appointees to competitive service positions. All other uses of a polygraph to screen applicants for and appointees to competitive positions are forbidden.

The regulations also set forth steps for determining whether agencies met the criteria of having a highly sensitive mission, and stipulated that approval to use the polygraph would be granted only for 1-year periods. Agencies intending to use the polygraph for personnel screening were required to prepare regulations and directives meeting certain minimum standards. The minimum

standards included directives concerning the specific purposes for which the polygraph may be used, and directives that a person to be examined must be informed as far in advance as possible of the intent to use the polygraph and of the fact that refusal to consent to a polygraph examination will not be made a part of the person's personnel file.

Also in response to the House Government Operations Committee's 1965 report, DOD proposed, and in part undertook, an extensive polygraph research program. And in July 1965, DOD issued directive 5210.48 (177) to regulate the conduct of polygraph examinations and improve selection, training, and supervision of its polygraph operators. Some of the results of the DOD research program were later reported in a scientific journal (29), but other reliability and validity studies proposed were never carried out (183).

Between 1967 and 1973 a number of bills were introduced which would have either limited the questions that could have been asked or banned altogether polygraph use by Federal agencies (170). None of these bills was enacted.

## The 1970's

Ten years after the 1964 hearings, this same House Government Operations subcommittee conducted another review of polygraph use by Federal agencies (169). In 1974 hearings, the subcommittee found that the use of polygraphs in the Federal Government had declined substantially since 1963. In fiscal year 1973, a total of 6,946 examinations were conducted, including 3,081 by NSA. This compared to 19,796 in 1963, excluding NSA and CIA. The subcommittee also found that there was not much additional research on polygraph validity. The only federally funded studies conducted had been those reported by the DOD Joint Services Group (183), and these studies were considered by DOD to be inadequate for determining the validity and reliability of Federal polygraph testing.

In a 1976 report based partly on the 1974 hearings, the House Government Operations Committee concluded that "the nature of research undertaken, both federally and privately funded, and

---

the results therefrom, have done little to persuade the committee that polygraphs . . . have demonstrated either their validity or reliability in differentiating between truth and deception, other than possibly in a laboratory situation” (171). The 1976 report concurred with the 1965 report that “There is no ‘lie detector’ “ (171). Because of the polygraph’s “unproven technical validity” and the suggestion that the “inherent chilling effect on individuals subjected to such examination clearly outweighs any purported benefit to the investigative function of the agency, ” the Committee recommended a complete ban on the use of polygraphs by all Federal Government agencies for all purposes. However, 13 committee members dissented, asserting both that the hearings had been held during an entirely different Congress, and participated in by an entirely different group of Members, and that, while testimony at the hearings represented a wide diversity of views, no witness had urged prohibition of the polygraph for all purposes. The dissenters urged adoption of the recommendations originally proposed and voted on by the members who had participated in the hearings. These recommendations would have, in part, prohibited the use of polygraphs in all cases except “1) those clearly involving the Nation’s security, and 2) those in which agencies can demonstrate in compelling terms their need for use of such devices for their law enforcement purposes, and that such uses would not violate the fifth amendment or any other provision of the Constitution. ”

The concern with scientific validity and its implications for the Federal Government’s use of polygraph testing arose again in 1979 at hearings held on preemployment security clearance procedures by the House Permanent Select Committee on Intelligence, Subcommittee on Oversight (175). The subcommittee found that there had been insufficient research on the accuracy of the polygraph technique in screening job applicants and that “gaps in the statistics kept by the intelligence services do not make it possible to make the clear judgment that the polygraph is unique and indispensable” (173). The Director of Central Intelligence (DCI) was urged to conduct a study to validate the accuracy of the polygraph for preemployment screening. DCI did conduct a study in 1980

to examine the utility of polygraph tests, but it was not a validity study (165).

As shown in figure 1, in addition to interest in Federal use of polygraphs, Congress has shown interest in the use of polygraph examinations by private employers, in part because of constitutional and privacy issues (see, e.g., 169,172, 173; the Privacy Protection Study Commission Report (128) mandated by Public Law 93-579; and several laws introduced since 1967). Various congressional committees have questioned the validity of polygraph testing in a private employment context, in particular as a condition for employment. Nevertheless, attempts to enact Federal legislation regulating the use of polygraph examinations by private employers and/or the Federal Government have not been successful.

### **The 1980’s**

In the recent past, the executive branch has again taken initiatives concerning the Federal use of polygraph testing. In April 1982, a DOD select panel reviewed the DOD personnel security program (180) and expressed dissatisfaction because of inconsistency in polygraph use across component programs (as did the U.S. Congress (173)), and the lack of reinvestigations. The panel observed that military personnel, unlike civilians, were appointed to NSA and allowed access to Sensitive Compartmented Information (SCI) without undergoing a polygraph examination. In addition, personnel could continue to get clearances throughout their careers without ever being subject to reexamination. The DOD panel recommended a broadened application of the polygraph for security screening purposes, and selective use of counterintelligence scope polygraph examinations during periodic reinvestigations. The panel noted that the recommended expanded use of the polygraph would require changes in DOD Directive 5210.48.

On August 6, 1982, the Office of the Deputy Secretary of Defense (39) issued a memorandum requiring employees with SCI access to agree to submit to polygraph examinations on an aperiodic basis, and revised DOD Directive 5210,48 accordingly. Later in 1982 and again in early and mid-1983, further revisions to DOD Directive 5210.48

were drafted (181). In 1983, the President issued a National Security Decision Directive (NSDD-84) also authorizing broader use of the polygraph. Congress responded to these developments by conducting several sets of hearings, by requesting OTA and General Accounting Office studies, and by passing an amendment to the DOD appropriations authorization bill (S.675) putting a moratorium until April 15, 1984, on any revisions to DOD Directive 5210.48 retroactive to August 5, 1982. On October 19, 1983, DOJ announced a new administration polygraph policy that would permit further expansion in polygraph use. The DOD draft revisions, NSDD-84, and administration polygraph policy are discussed in more detail below.

### **Draft Revisions to DOD 5210.48**

The draft revisions to the DOD polygraph regulations have gone through several iterations. For the purposes of this validity study, a primary proposed revision (as of the March 1983 draft) is to authorize the use of the polygraph for determining initial and continuing eligibility of DOD civilian, military, and contractor personnel for access to highly classified information (SCI and/or special access). The use of the polygraph in determining continuing eligibility would be on an aperiodic (i. e., irregular) basis (MI).

Also, the proposed revisions provide that refusal to take a polygraph examination, when established as a requirement for selection or assignment or as a condition of access, may, after consideration of all other relevant factors, result in adverse consequences for the individual. Adverse consequences are defined to include non-election for assignment or employment, denial or revocation of clearance, or reassignment to a nonsensitive position.

Technically, these expanded uses of the polygraph are considered to be part of personnel security investigations. Use of the polygraph within DOD is already authorized under the existing 1975 version of 5210.48 for various criminal, counter-intelligence, and intelligence purposes,

A detailed review of the proposed changes is beyond the scope of this technical memorandum.

### **NSDD-84**

On March 11, 1983, the President issued a National Security Decision Directive intended, according to DOJ officials, to help safeguard against unlawful disclosure of properly classified information. One provision of NSDD-84 requires that persons with authorized access to classified information sign a nondisclosure agreement, and that persons with access to SCI must also agree to pre-publication review. These provisions are outside the scope of this memorandum, as is a full analysis of NSDD-84.

With respect to the polygraph, NSDD-84 in effect authorizes agencies and departments to require employees to take a polygraph examination in the course of internal investigations of unauthorized disclosures of classified examinations. NSDD-84 also provides that refusal to take a polygraph test may result in adverse consequences. NSDD-84 permits administrative sanctions, including denial of security clearance, to be applied even when a person is not subject to a criminal investigation (184).

### **Administration Polygraph Policy**

On October 19, 1983, DOJ announced a comprehensive administration policy on Federal agency polygraph use. The policy authorizes polygraph testing:

1. as a condition of initial or continuing employment with or assignment to agencies with highly sensitive responsibilities directly affecting national security;
2. as a condition of access to highly sensitive categories of classified information;
3. to investigate serious criminal cases; and
4. to investigate serious administrative misconduct cases including unauthorized disclosure of classified information (185a).

The policy in essence authorizes use of the polygraph on a Government-wide basis for the expanded polygraph uses proposed by DOD. Thus, for example, the policy provides agency heads with the authority to give polygraph examinations on a periodic or aperiodic basis to randomly selected employees with access to highly sensitive information, and to deny such access to employees refusing to take a polygraph exam.

## SCIENTIFIC VALIDITY AND POLYGRAPH RESEARCH REVIEWS

Thus, recent polygraph policy actions have renewed interest in and debate over the scientific validity of the polygraph. Reviews of scientific literature form the principal means to cumulate research findings and are especially important in order to assess the validity of polygraph testing. Single research studies, no matter how well conducted, cannot answer global questions about validity and must be considered in relation to other evidence. Both because research evidence about polygraph testing has rapidly increased, especially within the last 10 years, and because there have been disagreements about the nature of evidence about polygraph testing, there have been a number of such reviews. These reviews are important, because they are frequently cited in both legal and legislative considerations and because they serve to shape future research,

Underlying each of the reviews is the application of a set of criteria, only sometimes made explicit, regarding the validity of individual studies and their implications for overall assessments of polygraph testing accuracy. As introduction to the scientific reviews, the nature of these criteria is described. The reviews, themselves, are then summarized and a preliminary analysis of discrepancies among reviews is presented. More detailed analysis of individual validity studies is provided in chapters 4 and .5,

### Definitions of Scientific Validity

#### Validity

The validity of polygraph testing means, in nontechnical terms, accuracy of the test in detecting deception and truthfulness. The problem of assessing polygraph validity is especially difficult, not only because polygraph tests take a number of forms, but also because validity has different dimensions and can be measured in a number of ways. There are, as a result, a number of different forms of validity associated with polygraph examinations depending on the type of polygraph test as well as on its use (e. g., employee screening v. investigation of a criminal suspect). These difficulties underlie, in part, the failure to have

developed assessments of polygraph validity that are accepted by the scientific community.

In order to make explicit the criteria for validity used in this assessment, below are described several dimensions of validity and how they are assessed. This description is based both on standards for psychological/psychometric tests (cf. 3,5) and criteria to evaluate research designs (cf. 41,147). Although criteria for validity can be described objectively, it should be noted that it is essentially a qualitative judgment as to whether (or, to what extent) a given criterion is met. In addition, assessments of the "preponderance" of evidence necessary in order to assess the overall validity of polygraph testing are similarly subjective. In chapters 4 and 5, a systematic analysis of available research is attempted, although it should be recognized that there are a number of ways to conduct such evaluations, each of which may yield a somewhat different outcome.

#### Reliability

Assessment of any test's validity is based on the assumption that the test consistently measures the same properties. This consistency, known as reliability, is usually the degree to which a test yields repeatable results (i. e., the extent to which the same individual retested is scored similarly).

Reliability also refers to consistency across examiners/scorers. A reliable polygraph test should yield equivalent outcomes when subjects are retested and, as well, be scored similarly by individuals other than the initial examiner. For example, if a polygraph examiner reviewed a set of charts and concluded that a subject was deceptive, any other polygraph examiner should be able to review the same charts and conclude that deception was indicated. This illustrates interrater-reliability. Such reliability might be affected by the amount and type of training of examiners.

The present study focuses primarily on validity because if a testing procedure is not measuring what it purports to measure (validity), it matters little that it can measure the same thing again and again. Examiners who consistently agree that they are seeing "deception" may in fact be measur-

ing anxiety or some other form of arousal. Reliability is, however, a necessary condition for validity to be established. A test that is valid will, necessarily, be reliable.

### Construct Validity

Construct validity refers, in broad terms, to whether a test adequately measures the underlying trait it is designed to assess. A polygraph test is designed to detect deception. It is therefore important to clearly define the construct of deception, and distinguish it from other concepts such as guilt.

To measure construct validity, it is necessary to both describe the construct and show its relation to a conceptual framework. Construct validation, thus, requires that a test be based on some theory or conceptual model. Since different types of polygraph tests have different theoretical bases (see ch. 2), there are multiple forms of construct validity for the polygraph. Construct validity is established by various means. Most importantly, based on theoretical predictions of how items should interrelate or how other tests should intercorrelate, actual evidence (e. g., scores from similar tests) is examined. If no such predictions are possible, it is impossible to establish construct validity.

### Criterion Validity

Although from a theoretical point of view construct validity is most important, from a practical point-of-view, criterion validity is the central component of a validity analysis. This aspect of validity refers, in the case of polygraph examinations, to the relationship between test outcomes and a criterion of ground truth. In this respect, criterion validity is what is meant by test accuracy. In the absence of construct validity evidence, however, it is difficult to determine to what extent criterion validity data can be generalized. In some situations, it is not clear which aspects of a test are responsible for accuracy, and what factors cause a test to be inaccurate.

### Research Design

The above validity criteria are those which are typically assessed in considering evidence about

the usefulness of a test. A related set of validity criteria are also used to evaluate the validity of any single study design. These research design criteria include, most importantly, internal and external validity (cf. 41,147).

Internal validity refers to the degree to which a study has controlled for extraneous variables which may be related to the study outcome. External validity refers to the established generalizability of a study to particular subject populations and settings. Internal validity in the case of a study of polygraph testing is usually enhanced by the presence of control groups. Typically, such conditions of an experiment permit analysis of variables such as different question formats. In most field studies, internal validity is difficult to establish since the investigation cannot control or, in many cases, have definitive knowledge about whether a subject is guilty or innocent.

External validity is simply the nature of the subjects and settings tested. The broader the population examined and the type of setting investigated, the wider that study's results can be generalized. In a parallel way, the more similar the research situation to the "real life" situation, the greater a study's external validity. Evidence about external validity is developed both from investigations that test a broad range of subjects and situations and from investigations that identify subject and setting interactions with polygraph test outcomes. The broader the population examined and the type of setting or the more similar it is to the situation for which one wants to use a test or a theoretical construct, the greater a study's external validity.

### False Positives and Negatives

With any test, the possibility exists of false positives and negatives. False positives are decisions that individuals are being deceptive when they are providing truthful responses. Their charts are scored as showing a "deceptive" reaction for some other reason. False negatives are decisions that individuals are not being deceptive when in fact they are being deceptive. There are a number of reasons why such false outcomes might be obtained and, in part, they depend on the criteria (e. g., amount of physiological change) used to indicate deception or truthfulness.

The rate of false positives or negatives is sometimes difficult to establish because, in research studies, a number of criteria for deception/nondeception may be applied. Thus, for example, in studies which employ numerical scoring for polygraph charts, depending on the scoring system (e. g., cutoff points), different diagnoses will be made. The rate of false positives and negatives may also depend on the examiner's perception of the "base rate" of guilt/innocence.

In some cases, the examiner will deal mostly with deceptive subjects (e. g., in certain criminal investigation contexts) and, thus, may be predisposed to make false positive diagnoses. In other settings (e. g., some personnel screenings), an examiner may test only a small number of deceptive subjects and, then, may be predisposed to false negative decisions. Regardless of rates, assessment of conditions that contribute to either type of error is a focus of the research literature.

### Reviews of Polygraph Validity

Since at least 1973, a number of polygraph researchers and psychologists interested in physiological detection of deception have reviewed available scientific literature to assess the validity and reliability of polygraph testing. Most such reviews focus on studies of criterion validity, although a growing number of investigations deal with construct validity. The most important difference among these criterion studies has to do with whether they are conducted in actual field situations or in "analog" situations.

#### Field Studies

For purposes of this technical memorandum, field studies are those studies or "naturally" occurring polygraph test situations; i.e., studies in which the researcher does not exercise experimental control over the situation in which the crime or other event occurred. Not exercising experimental control means that the researcher does not systematically assign people to conditions of, for example, guilt or innocence. We refer hereto "field" studies but others (e.g., 7) use the terminology "real" cases (v. "laboratory"). Abrams (1) differentiates between the laboratory and "actual criminal cases. "

In polygraph field studies, polygraph examiners' decisions are compared against some post hoc determination of whether suspects are guilty or innocent; i.e., "ground truth. " These post hoc determinations may, in different studies, consist of confessions by the presumably guilty party, decisions by a panel of attorneys or judges assembled specifically for a particular study who base their decisions on investigative files excluding references to polygraph decisions, judicial outcomes (dismissals, acquittals, convictions), as well as other criteria. The fact that determinations of guilt or innocence are made post hoc makes drawing conclusions from field studies difficult (126). In real life situations, truth is seldom available (62).

Attempts to use confessions, panel judgments, judicial outcomes, and other criteria as indicators of truth have their own problems. Individuals may confess to crimes which they did not commit (108). In addition, individuals are sometimes falsely convicted (34). Panel decisions may be generalizable only to cases in which sufficient investigative information is available to make a decision without the addition of polygraph testing. One can never be certain that the panel decision is indeed correct, and the panel and the polygraph examiner may have been exposed to the same prior information (62). Thus, while field studies provide the most direct evidence about polygraph test validity, they have been criticized because they do not adequately meet the standards of "ground truth" to establish criterion validity.

#### Comparison of Reviews

A number of independent reviews (listed in table 2) of the field evidence on polygraph testing were assessed in order to determine reasons for differences among reviews. The reviews differ in a number of respects. In part, reviewers' conclusions differ because they include different kinds of studies and even different studies (despite, in several cases, having had the same studies available to them). In addition, some reviews differentiate between accuracy in detecting deceptive v. nondeceptive subjects, emphasizing the problems of false positives and false negatives; others aggregated the overall accuracy rates across both groups of subjects. Finally, there are differences

Table 2.—Reviews of Field Studies of Polygraph Validity

Study <sup>b</sup> /N <sup>e</sup>	Year	Q <sup>c</sup>	D <sup>d</sup>	Abrams (1973)	Horvath (1976)	Maskin & Podlesny (1979) <sup>a</sup>	Lykken (1981)	Ben-Shakhar (1982)	Ansley (1983) <sup>i</sup>
Bitterman and Marcuse (N = 2Ex, 8IS)	1947	R/I	A	[100]	NR	NR	EX <sup>u</sup>	NR	[100]
Ben-Ishai (N = 1E, 100C)	1962	C	A	NR <sup>h</sup>	NR	NR	NR	NR	[97.0] <sup>j</sup>
B (N = E, 10C)	unpublished	C	A	NR	NR	NR	NR	NR	[100]
Bersh (N = 4E, 216C)	1962	C	G	NR	NR	NR	NR	NR	NR
Bersh (N = 4E, 216C)	unpublished	C	G	NR	NR	NR	NR	NR	NR
Bersh (N = 4E, 216C)	1969	C	G	NR	NR	NR	NR	NR	NR
Horvath and Reid	97	S	A	[79.1-91.4] <sup>k</sup>	[74.6-92.4] <sup>l</sup>	75-89 <sup>k</sup>	75-92 <sup>l</sup>	85.6	[96.2] <sup>l</sup>
Hunter and Ash	1973	C	G	—	EX <sup>l</sup>	89-94 <sup>k</sup>	EX <sup>l</sup>	89.3	NR
Hunter and Ash	1973	C	G	—	EX <sup>l</sup>	88	EX <sup>l</sup>	90.5	NR
Wicklander and Hunter	1975	S	G	—	EX <sup>l</sup>	86	EX <sup>l</sup>	87.1	NR
Wicklander and Hunter	1975	S	G	—	EX <sup>l</sup>	86	EX <sup>l</sup>	85.7	NR
Wicklander and Hunter	1975	S	G	—	EX <sup>l</sup>	93	EX <sup>l</sup>	90.0	NR
Wicklander and Hunter	1975	S	G	—	EX <sup>l</sup>	85	EX <sup>l</sup>	86.7	NR
Wicklander and Hunter	1975	S	G	—	EX <sup>l</sup>	85	EX <sup>l</sup>	83.8	NR
Wicklander and Hunter	1975	S	G	—	EX <sup>l</sup>	93	EX <sup>l</sup>	90.5	NR
Raskin (3 studies)	1988 <sup>g</sup>	L	G	—	—	93-100 <sup>m</sup>	NR	NR	NR
Raskin (3 studies)	1988 <sup>g</sup>	L	G	—	—	69-95 <sup>m</sup>	NR	NR	NR
Barland and Raskin (N = 102C)	976	C	G	—	—	NR	98	83.0	NR
Barland and Raskin (N = 102C)	976	C	G	—	—	NR	45	29.4	NR
Horvath	1977	C	A	—	[78.7-89.7] <sup>o</sup>	NR	77	77.1	NR
Horvath	1977	C	A	—	EX <sup>l</sup>	NR	51	51.1	NR
Davidson	1979	C	I	—	—	—	NR	NR	EX <sup>l</sup>
Edwards (N = 80E, 959C)	1981	C	A	—	—	—	NR	NR	[98.3] <sup>p</sup>
Elaad (N = 184C)	unpublished	C	A	—	—	—	NR	NR	NR
Elaad (N = 184C)	1976	C	A	—	—	—	NR	NR	NR
Peters (N = 172C)	unpublished	C	A	—	—	—	NR	NR	[98.3] <sup>q</sup>
Peters (N = 172C)	1982	C	A	—	—	—	NR	NR	[95.1] <sup>r</sup>
Widacki (N = 38C)	unpublished	Z	A	—	—	—	—	—	[95.8] <sup>r</sup>
Widacki (N = 38C)	1982	Z	A	—	—	—	—	—	[95.8] <sup>r</sup>
Author's conclusion (if any)				98††	NR	90	54.7 <sup>s</sup>	83.84 <sup>t</sup>	97.6
Author's conclusion (if any)				98††	NR	89	54.7 <sup>s</sup>	76.81 <sup>t</sup>	97.6

<sup>a</sup>Not technically a review. Original review (126) relied on lab evidence. When criticized by Lykken, authors responded in part by citing field studies.

<sup>b</sup>Presented in chronological order.

<sup>c</sup>Q = question type (C = control question; R/I = relevant/irrelevant; Z = zone of comparison).

<sup>d</sup>D = correct decision or rate of agreement (i.e., hit rate) for guilty (deceptive) (G) or innocent (nondeceptive) (I) subjects.

<sup>e</sup>A = average hit rate (in cases where G and I are not reported separately).

<sup>f</sup>N = number. Ex = number of examiners; E = number of evaluators; C = number of cases evaluated; S = number of subjects when actual subjects are tested in the study.

<sup>g</sup>It is important to note Ansley's method of calculating overall accuracy. When there are inconclusives (or disagreements between the original examiners and the evaluator), Ansley divides the percent inconclusive by 2 and adds one-half the percentage to the accuracy percent reported by the original researcher.

<sup>h</sup>The 74.6 percent figure is for majority decision; 92.4 percent for unanimous decision.

<sup>i</sup>The lower rate is the accuracy rate for inexperienced polygraphers; the higher rate is the accuracy rate for experienced polygraphers.

<sup>j</sup>Excluded because they are seen as reliability studies. Horvath reasons that they used confessions as a criterion and because confessions are not independent of polygraphers' examinations: "their usefulness as a criterion measure for estimating validity is limited." Lykken excludes them because both examiners and evaluators were Reid polygraphers.

<sup>k</sup>Lower rate is for nonnumerical evaluation; higher rate for numerical evaluations.

<sup>l</sup>Focus of study was to test the "friendly polygrapher" hypothesis (see Raskin, Barland, and Podlesny, 1978).

<sup>m</sup>78.7 percent = agreement between simple majority on panel and polygraph decision; 86.7 percent = agreement between 4/5 majority on panel and polygraph; 89.7 percent = agreement between judicial outcome and polygraph (where judiciary not aware of polygraph results).

<sup>n</sup>Of 2,433 polygraph examinations in the original pool, only 959 (39.4 percent) could be verified as to truthfulness or not. Of these 959, 943 were found to be correct. The 98.3 percent figure reported by Ansley represents 943/959 and does not take into account the entire pool of examinations.

<sup>o</sup>Actually, of 1984 cases that were confirmed, 10 were inconclusive and 6 were errors. The percent accuracy given in Ansley's review excludes the inconclusive results and adds half the error rate to the accuracy rate. If one appropriately excludes both and inconclusives and the error rate, the accuracy rate is at most 87 percent.

<sup>p</sup>The lower rate is for studies using confessions criteria; higher rate is for studies using panel of experts criteria. Figures are rounded.

<sup>q</sup>Use control test only.

<sup>r</sup>Calculations based on Lykken's conclusion that 6 out of 11 innocent subjects were judged deceptive.

<sup>s</sup>††Abrams uses many other studies to come to this conclusion. Many are anecdotal. Other researchers (e.g., Horvath) say they should not be relied on.

in the way accuracy rates were calculated, in particular, how inconclusive are handled. Each of these differences has important implications for the conclusions developed by the reviews,

Several reviews (1,81) conducted 5 to 10 years ago reported relatively positive conclusions based on an evaluation of the scientific literature.

Abrams (1) in 1973 reviewed reports of the polygraph's accuracy dating from 1917, including anecdotal as well as experimental data. He calculated approximate estimates of overall accuracy from this data, noting, however, that "it is almost meaningless to total and average these findings because of the great discrepancy in experimental paradigms and the instruments employed." He reported that in studies with complete verification of ground truth, diagnoses were 100 percent correct. In other field studies prior to 1963 Abrams calculated an accuracy rate of 98 percent. In laboratory experiments prior to 1963, Abrams estimated the average accuracy rate of 81 percent. Averaging the results of the reports between 1963 and 1973, Abrams estimate of laboratory and field research accuracy was 83 and 98 percent, respectively. Horvath's (6) review in 1976 used somewhat more stringent criteria in selecting data than did Abrams. His review does not include an overall average accuracy rate calculated across studies.

The early positive views of the polygraph's worth have recently been challenged by Lykken (108) and, to some extent, by Ben-Shakhar, et al. (28). Lykken in 1981 challenged the theoretical assumptions of the most prevalent question technique, the control question technique (CQT), and asserted that an average 50-percent false positive rate supported his theoretical challenge. Lykken, however, continues to believe that particular polygraph techniques are useful (i.e., the detection of guilt by measuring physiological arousal) and offers the use of the guilty knowledge technique as a way to increase overall validity. Adoption of Lykken's suggestion would preclude the use of the polygraph for preemployment testing and periodic checking.

Ben-Shakhar, et al.'s (28), analysis also limited their assessment of the polygraph to CQT. Their 1982 assessment of existing polygraph field research indicated that polygraph testing was 83 to

84 percent accurate for guilty suspects and 76 to 81 percent accurate for innocent suspects. As a result, Ben-Shakhar, et al., concluded that examiners tend to value detection of guilty suspects highly, even at the risk of falsely classifying innocent suspects; their conclusion concurs with Lykken's. Ben-Shakhar, et al., in conducting their review, employ a utility theory approach based on Bayes' theorem. They predict dramatically different utility rates based on different base rate assumptions.

Although these recent reviews, by authors who are not professional polygraphers, cast doubt on the validity of at least the most common polygraph technique, a more recent review by Ansley (7) comes to the most positive conclusions since those of Abrams. Ansley's 1983 review is an important review because it represents the views of NSA's, chief polygraph examiner. (NSA conducts the largest number of polygraph examinations of any Federal agency, ) As shown in table 2, Ansley concludes that field research shows a 97.2-percent validity rate and laboratory research a 93.2-percent validity rate. Based on these validity calculations as well as separate calculations for reliability and utility, Ansley concludes that the polygraph is "clearly an excellent adjunct to the selection process."

Unfortunately, for the most part, polygraph reviews contained in table 2 do not explicitly state their study selection criteria (see 63). The result is that a number of different studies have been included in various reviews, each of which presents different problems for interpretations of validity. The kinds of studies include reports of single criminal investigations in which the actual solution to the crime is the criterion for validity; studies in which "blind" polygraph interpreters compare their polygraph chart evaluations to "ground truth" as established by confession; and studies in which the judgment of legal professionals, actual judicial outcome, or in one case, the judgment of a single psychologist, is used to establish ground truth.

Some reviews do specify criteria for exclusion. Lykken, for example, does not include studies of single criminal investigations. Abrams, on the other hand, includes in his review a number of

such studies (e. g., 30, 103). Lykken's reasoning was that in single criminal investigations, the examiner has a large chance of being accurate (depending on the number of suspects) merely by calling everyone innocent. The fact that other reviewers do not include Bitterman and Marcuse and other such reports, implies that they accept and other such reports, implies that they accept and other such reports, implies that they accept graph testing validity, particularly in view of the Lykken's evaluation of the usefulness of such studies as indicators of validity. It is possible that false negative decisions. If, for example, the results of such reports could be useful in assessing innocent correct rate is 80 percent but the remaining polygraph screening of large numbers of individuals in specific incident cases, such as might be the case in unauthorized disclosure investigations. However, additional factors limit the external validity of Bitterman and Marcuse and other such studies. In Bitterman and Marcuse, for example, the investigators were psychology professors apparently conducting their first polygraph tests, and they did not use accepted polygraph procedures or instruments. There are no recent systematic studies of specific incident investigations involving a large number of suspects.

There is strong disagreement among reviewers about whether another group of studies should be included as indicators of validity. These studies were conducted with records selected from the files of the John E. Reid & Associates polygraph firm. A group of cases was used which the authors considered to be "verified" by confession of the guilty suspect (in most cases they were also verified by some form of corroboration (37)). The polygraph charts in these cases are then reinterpreted by a group of polygraphers who are "blind" to (i. e., do not know) the suspect's guilt or innocence. The degree of agreement of the "blind" evaluators to verify guilt or innocence is the test of validity. Two reviewers (Horvath, Lykken) explicitly excluded the group of studies conducted based on Reid files. Horvath excluded them because they used confessions as a criterion (confessions not being independent of the polygraph examinations), and Lykken because both examiners and "blind" evaluators were polygraphers from the same firm. His claim was that the studies were, thus, "merely demonstrations that Reid's examiners score charts in a similar way" (108) and so were estimates of reliability rather than validity. However, reviews by Raskin and Podlesny (138) and Ben-Shakhar, et al. (27), each use all four Reid studies to assess validity.

Conclusions about the validity of the polygraph may depend on whether the reviewer attends to the average accuracy rate or to the accuracy for guilty and innocent subjects separately. The fact that other decision statistics contributes to the ability to make an accurate assessment of polygraph testing validity, particularly in view of the concern over both high false positive and high false negative decisions. If, for example, the innocent correct rate is 80 percent but the remaining 20 percent consists of inconclusive" or of false negatives. In some cases (e. g., preemployment screening), inaccurately designating nondeceptive people as deceptive may have worse consequences for the employee than inaccurately deciding that deceptive individuals are nondeceptive. In some cases (e. g., a heinous crime by a potential repeat offender, infiltration by a foreign agent), a false negative may have serious consequences.

In only two reviews (Ben-Shakhar, Lykken) are summary percentages provided in terms of the percent accurately detected for both guilty and innocent; in other reviews, these figures are presented as the average percent of accurate detections. In some cases, the percent inaccurately "detected" as nondeceptive (when they were really deceptive) or deceptive (when they were really nondeceptive) as well as percent inconclusive were also reported by reviewers. But for purposes of clarity these have been omitted from table 2.

Another reason reviews differ about the results of the same studies is the fact that they make different decisions about the base rate of subjects or cases that are included. If, for example, a panel cannot make a decision about 30 percent of the cases (e. g., 22), some reviewers will omit the number of nonagreements from the number included in the accuracy rate and base accuracy percentages on only the remaining cases. This accounts for the difference between Horvath and Ben-Shakhar, et al., analyses of the Barland and Raskin results. In other studies (and reviews of those studies, e.g., Ansley, Abrams) inconclusive polygraph results are excluded from the analysis. This has the effect of inflating the accuracy rates.

Apart from the different base rates on which most of the reviewers calculated accuracy rates (see above), one source of different accuracy rates applies uniquely to Ansley (7). In any case in which there is not 100-percent accuracy, the Ansley review computes validity by dividing the difference between the accuracy rate and 100 percent (the so-called error rate) in half and adds half of the difference to the accuracy rate. Ansley uses this procedure on the grounds that on the basis of chance, errors were probably half in favor of the panel (or other criterion measure) and half in

favor of the examiners. For example, in the Bersh study, half of the difference between the typically reported 92.4-percent rate and 100 percent *is* 7.6 which Ansley divides in half, leaving a validity rate of 96.2 and an error rate of 3.8 percent. The same method is used for the Peters, Elaad, and Widacki studies, for which the preadjustment validity rates are 90.2, 96.6, and 91.6 percent, respectively. Each of these studies, particularly Elaad (see ch. 4), have other problems of interpretation as well,

## CONCLUSIONS

Central to legal, legislative, and scientific assessment of polygraph tests are their validity. Yet, despite many decades of judicial, legislative, and scientific discussion, no consensus has emerged about the accuracy of polygraph tests. One explanation is that scientific criteria for validity deal with a number of dimensions and that the criteria vary widely among specific research studies. In order to assess overall polygraph examination validity, it will be necessary to examine details of each of the relevant studies. Such analysis is presented in chapters 4 and 5.

Another explanation is that polygraph testing has been viewed as a single technique. Thus, despite testimony (e.g., 137) which urged differential consideration of polygraphs used in, for example, employment screening and criminal investigations, the scientific evidence for particular purposes has not been differentiated. As is demonstrated by the analysis of scientific literature (here and in chs. 5 and 6), in assessing validity it is necessary to separate clearly the purposes for which polygraph examinations are conducted and the types of techniques employed.