Executive Summary

For as long as human beings have deceived one another, people have tried to develop techniques for detecting deception and finding truth. Lie detection took on aspects of modern science with the development in the 20th century of techniques intended for the psychophysiological detection of deception, most prominently, polygraph testing. The polygraph instrument measures several physiological processes (e.g., heart rate) and changes in those processes. From the charts of those measures in response to questions on a polygraph test, sometimes aided by observations during the polygraph examination, examiners infer a psychological state, namely, whether a person is telling the truth or lying.

Polygraph testing is used for three main purposes: event-specific investigations (e.g., after a crime); employee screening, and preemployment screening. The different uses involve the search for different kinds of information and have different implications. A question asked about a specific incident (e.g., "Did you see the victim on Monday" or "Did you take the file home yesterday?") often has little ambiguity, so it is clear what facts provide the criterion for a truthful answer.

For employee screening, there is no specific event being investigated, and the questions must be generic (e.g., "Did you ever reveal classified information to an unauthorized person?"). Both examinee and examiner may have difficulty knowing whether an answer to such a question is truthful unless there are clear and consistent criteria that specify what activities justify a "yes" answer. Examinees may believe they are lying when providing factually truthful responses, or vice versa. Polygraph tests might elicit admissions to acts not central to the intent of the question and these answers might be judged either as successes or failures of the test. In this regard, we have seen no indication of a clear and stable agreement on criteria for judging answers to security screening polygraph questions in any agency using them.

The use of polygraph testing for preemployment screening is even more complicated because it involves inferences about future behavior on the basis of information about past behaviors that may be quite different (e.g., does past use of illegal drugs, or lying about such use on a polygraph test, predict future spying?).

The committee's charge was specifically "to conduct a scientific review of the research on polygraph examinations that pertains to their validity and reliability, in particular for personnel security screening," that is, for the second and third purposes. We have focused mainly on validity because a test that is reliable (i.e., produces consistent outcomes) has little use unless it is also valid (i.e., measures what it is supposed to measure). Virtually all the available scientific evidence on polygraph test validity comes from studies of specific-event investigations, so the committee had to rely heavily on that evidence, in addition to the few available studies that are relevant for screening. The general quality of the evidence for judging polygraph validity is relatively low: the substantial majority of the studies most relevant for this purpose were below the quality level typically needed for funding by the National Science Foundation or the National Institutes of Health.

SCIENTIFIC EVIDENCE

Basic Science

Almost a century of research in scientific psychology and physiology provides little basis for the expectation that a polygraph test could have extremely high accuracy. Although psychological states often associated with deception (e.g., fear of being judged deceptive) do tend to affect the physiological responses that the polygraph measures, these same states can arise in the absence of deception. Moreover, many other psychological and physiological factors (e.g., anxiety about being tested) also affect those responses. Such phenomena make polygraph testing intrinsically susceptible to producing erroneous results. This inherent ambiguity of the physiological measures used in the polygraph suggest that further investments in improving polygraph technique and interpretation will bring only modest improvements in accuracy.

Polygraph research has not developed and tested theories of the underlying factors that produce the observed responses. Factors other than truthfulness that affect the physiological responses being measured can vary substantially across settings in which polygraph tests are used. There is little knowledge about how much these factors influence the outcomes of polygraph tests in field settings. For example, there is evidence suggesting that truthful members of socially stigmatized groups and truthful examinees who are believed to be guilty or believed to have a high likelihood of being guilty may show emotional and physiological responses in polygraph test situations that mimic the responses that are expected of deceptive individuals. The lack of understanding of the processes that underlie polygraph responses makes it very difficult to generalize from the results obtained in specific research settings or with particular subject populations to other settings or populations, or from laboratory research studies to real-world applications.

Evidence on Polygraph Accuracy

Scientific evidence relevant to the accuracy of polygraph tests for employee or preemployment screening is extremely limited. Only one field study, which is flawed, provides evidence directly relevant to accuracy for preemployment screening. A few additional laboratory studies are relevant to preemployment or employee screening, but they are more analogous to specific-incident investigations than to screening because the deceptive examinee is given a precise recent incident about which to lie.

Of the 57 studies the committee used to quantify the accuracy of polygraph testing, all involved specific incidents, typically mock crimes (four studies simulated screening in the sense that the incidents were followed by generic screening-type questions). The quality of the studies varies considerably, but falls far short of what is desirable. Laboratory studies suffer from lack of realism, and in the randomized controlled studies focused on specific incidents using mock crimes, the consequences associated with lying or being judged deceptive almost never mirror the seriousness of these actions in real-world settings in which the polygraph is used. Field studies have major problems with identifying the truth against which test results should be judged. In addition, they suffer from problems associated with heterogeneity and lack of control of extraneous factors and more generally, they have lower quality than could be achieved with careful study design. Moreover, most of the research, in both the laboratory and in the field,

does not fully address key potential threats to validity. For these reasons, study results cannot be expected to generalize to practical contexts.

Estimates of accuracy from these 57 studies are almost certainly higher than actual polygraph accuracy of specific-incident testing in the field. Laboratory studies tend to overestimate accuracy because laboratory conditions involve much less variation in test implementation, in the characteristics of examinees, and in the nature and context of investigations than arise in typical field applications. Observational studies of polygraph testing in the field are plagued by selection and measurement biases, such as the inclusion of tests carried out by examiners with knowledge of the evidence and of cases whose outcomes are affected by the examination. In addition, they frequently lack a clear and independent determination of truth. Due to these inherent biases, observational field studies are also highly likely to overestimate real-world polygraph accuracy.

> **CONCLUSION: Notwithstanding the limitations of the quality of the empirical research and the limited ability to generalize to real-world settings, we conclude that in populations of examinees such as those represented in the polygraph research literature, untrained in countermeasures, specific-incident polygraph tests can discriminate lying from truth telling at rates well above chance, though well below perfection. Because the studies of acceptable quality all focus on specific incidents, generalization from them to uses for screening is not justified. Because actual screening applications involve considerably more ambiguity for the examinee and in determining truth than arises in specific-incident studies, polygraph accuracy for screening purposes is almost certainly lower than what can be achieved by specific-incident polygraph tests in the field.**

The accuracy levels in the four screening simulations in our sample, which include a validation study of the Test for Espionage and Sabotage (TES) used in the employee security screening program of the U.S. Department of Energy (DOE), are in the range reported for other specific-incident laboratory studies. The one field study of actual screening presents results consistent with the expectation that polygraph accuracy in true screening situations is lower.

Countermeasures

Countermeasures pose a potentially serious threat to the performance of polygraph testing because all the physiological indicators measured by the polygraph can be altered by conscious efforts through cognitive or physical means. Certain countermeasures apparently can, under some laboratory conditions, enable a deceptive individual to appear nondeceptive and avoid detection by an examiner. It is unknown whether a deceptive individual can produce responses that mimic the physiological responses of a nondeceptive individual well enough to fool an examiner trained to look for behavioral and physiological signatures of countermeasures. The available research provides no information on whether innocent examinees can increase their chances of achieving nondeceptive outcomes by using countermeasures. (It is possible that classified information exists on these topics; however, this committee was not provided access to such information and cannot verify its existence or relevance.)

CONCLUSION: Basic science and polygraph research give reason for concern that polygraph test accuracy may be degraded by countermeasures, particularly when used by major security threats who have a strong incentive and sufficient resources to use them effectively. If these measures are effective, they could seriously undermine any value of polygraph security screening.

POLYGRAPH USE FOR SECURITY SCREENING

The proportion of spies, terrorists, and other major national security threats among the employees subject to polygraph testing in the DOE laboratories and similar federal sites presumably is extremely low. Screening in populations with very low rates of the target transgressions (e.g., less than 1 in 1,000) requires diagnostics of extremely high accuracy, well beyond what can be expected from polygraph testing. Table S-1 illustrates the unpleasant tradeoffs facing policy makers who use a screening technique in a hypothetical population of 10,000 government employees that includes 10 spies, even when an accuracy is assumed that is greater than can be expected of polygraph testing on the basis of available research. If the test were set sensitively enough to detect about 80 percent or more of deceivers, about 1,606 employees or more would be expected "fail" the test; further investigation would be needed to separate the 8 spies from the 1,598 loyal employees caught in the screen. If the test were set to reduce the numbers of false alarms (loyal employees who "fail" the test) to about 40 of 9,990, it would correctly classify over 99.5 percent of the examinees, but among the errors would be 8 of the 10 hypothetical spies, who could be expected to "pass" the test and so would be free to cause damage.

Available evidence indicates that polygraph testing as currently used has extremely serious limitations in such screening applications, if the intent is both to identify security risks and protect valued employees. Given its level of accuracy, achieving a high probability of identifying individuals who pose major security risks in a population with a very low proportion of such individuals would require setting the test to be so sensitive that hundreds, or even thousands, of innocent individuals would be implicated for every major security violator correctly identified. The only way to be certain to limit the frequency of "false positives" is to administer the test in a manner that would almost certainly severely limit the proportion of serious transgressors identified.

CONCLUSION: Polygraph testing yields an unacceptable choice for DOE employee security screening between too many loyal employees falsely judged deceptive and too many major security threats left undetected. Its accuracy in distinguishing actual or potential security violators from innocent test takers is insufficient to justify reliance on its use in employee security screening in federal agencies.

Polygraph screening may be useful for achieving such objectives as deterring security violations, increasing the frequency of admissions of such violations, deterring employment applications from potentially poor security risks, and increasing public confidence in national security organizations. On the basis of field reports and indirect scientific evidence, we believe that polygraph testing is likely to have some utility for such purposes. Such utility derives from beliefs about the procedure's validity, which are distinct from actual validity or accuracy.

Polygraph screening programs that yield only a small percentage of positive test results, such as those in use at DOE and some other federal agencies, might be useful for deterrence, eliciting admissions, and related purposes. However, in populations with very low base rates of the target transgressions they should not be counted on for detection: they will not detect more than a small proportion of major security violators who do not admit their actions.

We have thought hard about how to advise government agencies on whether or how to use information from a diagnostic screening test that has these serious limitations. We note that in medicine, such imperfect diagnostics are often used for screening, though only occasionally in populations with very low base rates of the target condition. When this is done, either the test is far more accurate than polygraph testing appears to be, or there is a more accurate (though generally more invasive or expensive) follow-up test that can be used when the screening test gives a positive result. Such a follow-up test does not exist for the polygraph. The medical analogy and this difference between medical and security screening underline the wisdom in contexts like that of employee security screening in the DOE laboratories of using positive polygraph screening results—if polygraph screening is to be used at all—only as triggers for detailed follow-up investigation, not as a basis for personnel action. It also underlines the need to pay close attention to the implications of false negative test results, especially if tests are used that yield a low proportion of positive results.

A belief that polygraph testing is highly accurate probably enhances its utility for such objectives as deterrence. However, overconfidence in the polygraph—a belief in its accuracy that goes beyond what is justified by the evidence—also presents a danger to national security objectives. Overconfidence in polygraph screening can create a false sense of security among policy makers, employees in sensitive positions, and the general public that may in turn lead to inappropriate relaxation of other methods of ensuring security, such as periodic security re-investigation and vigilance about potential security violations in facilities that use the polygraph for employee security screening. It can waste public resources by devoting to the polygraph funds and energy that would be better spent on alternative procedures. It can lead to unnecessary loss of competent or highly skilled individuals in security organizations because of suspicions cast on them by false positive polygraph exams or because of their fear of such prospects. And it can lead to credible claims that agencies that use polygraphs are infringing civil liberties for insufficient benefits to the national security. Thus, policy makers should consider each application of polygraph testing in the larger context of its various costs and benefits.

ALTERNATIVES AND ENHANCEMENTS TO THE POLYGRAPH

CONCLUSION: Some potential alternatives to the polygraph show promise, but none has yet been shown to outperform the polygraph. None shows any promise of supplanting the polygraph for screening purposes in the near term.

The polygraph is only one of many possible techniques for identifying national security risks among federal employees. Other techniques attempt to detect deception from facial expressions, voice quality, and other aspects of demeanor; from measurements of brain activity and other physiological indicators; and from background investigations or questionnaires. Computerized analysis of polygraph records has the potential to improve the accuracy of test results by using more information from in polygraph records than is used in traditional scoring methods. This potential has yet to be realized, however, either in research or in practice.

We considered the potential to increase the capability to identify security risks by combining polygraph information with information from other screening techniques, for example, in serial screening protocols such as are used in medical diagnosis. There are good theoretical reasons to think appropriate procedures of this sort would improve detection of deception, but we found no serious investigations of such multicomponent screening approaches.

RESEARCH RECOMMENDATIONS

There has been no serious effort in the U.S. government to develop the scientific basis for the psychophysiological detection of deception by any technique, even though criticisms of the scientific grounding of polygraph testing have been raised prominently for decades. Given the heavy reliance of government on the polygraph, especially for screening for espionage and sabotage, the lack of a serious investment in such research is striking.

The limitations of the polygraph, especially for security screening, justify efforts to look more broadly for effective tools for deterring and detecting security violations. These might include modifications in the overall security strategies used in federal agencies, such as have been recommended by the Hamre Commission for DOE, as well as improved techniques for deterring and detecting security violations focused on individuals. Research offers one promising strategy for developing the needed tools.

We recommend an expanded research effort directed at methods for detecting and deterring major security threats, including efforts to improve techniques for security screening.

This effort should pursue two major objectives: (1) to provide federal agencies with methods of the highest possible scientific validity for protecting national security by deterring and detecting major security threats; and (2) to make these agencies fully aware of the strengths and limitations of the techniques they use. If the government continues to rely heavily on the polygraph in the national security arena, some of this research effort should be devoted to developing scientific knowledge that could put the polygraph on a firmer scientific foundation, develop alternative methods, or develop effective ways to combine techniques and methods. National security is best served by a broad research program on detecting and deterring security threats, not a narrow focus on polygraph research.

The research program should be open to supporting alternative ways of looking at the problems of deterrence and detection because there is no single research approach that clearly holds the most promise for meeting national security objectives. Thus, it might support research ranging from very basic work on fundamental psychological, physiological, social, and political processes related to deterring and detecting security threats to applied studies on implementing scientifically rooted methods in practical situations.

A substantial portion of our recommended expanded research program should be administered by an organization or organizations with no operational responsibility for detecting deception and no institutional commitment to using or training practitioners of a particular technique. The research program should follow accepted standards for scientific research, use rules and procedures designed to eliminate biases that might influence the findings, and operate under normal rules

of scientific freedom and openness to the extent possible while protecting national security.

The mandate should be broad and should include both basic and applied research. The program should use standard scientific advisory and decision-making procedures, including external peer review of proposals, and should support research that is conducted and reviewed openly in the manner of other scientific research. Classified and restricted research should be limited only to matters of identifiable national security. Mission agencies might well continue to conduct implementation-focused research on detecting deception, but their work should be integrated with the broader research program proposed here.

TABLE S-1 Expected Results of a Polygraph Test Procedure with an Accuracy Index of 0.90 in a Hypothetical Population of 10,000 Examinees that Includes 10 Spies

a: If detection threshold is set to detect the great majority (80 percent) of Spies

	Examinee's true condition		
Test result	Spy	Nonspy	Total
"Fail" test	8	1,598	1,606
"Pass" test	2	8,392	8,394
TOTAL	10	9,990	10,000

b: If detection threshold is set to greatly reduce false positive results

	Examinee's true condition		
Test result	Spy	Nonspy	Total
"Fail" test	2	39	41
"Pass" test	8	9,951	9,959
TOTAL	10	9,990	10,000