

Integration of Pre-Employment Polygraph Screening into the Police Selection Process

Mark Handler¹, Charles R. Honts², Donald J. Krapohl³,
Raymond Nelson⁴, and Stephen Griffin⁵

Abstract

The authors provide a polygraph primer for police psychologists involved in law enforcement personnel selection. Law-enforcement pre-employment polygraph examinations are a decision-support tool intended to add incremental validity to the personnel selection process. Problems stemming from the use of the polygraph may be related to misunderstanding of the polygraph test and to field practices surrounding the use of polygraphy in the police selection process. Potential problems can result from ineffective selection of test issues, poorly constructed test questions and misguided policies surrounding the use of the polygraph. The authors review polygraph screening history, research, and field practices, and suggest that using polygraph results alone to disqualify a candidate from employment is a misguided field practice. Suggestions are offered for maximizing the decision-support value of the polygraph. Polygraph examination targets are discussed, with emphasis on selecting actuarially derived predictors associated with increased success in law enforcement training and job performance. The authors provide recommendations for field practice, and propose that police psychologists may be most suited to effectively integrate the polygraph results and information into the hiring recommendation process.

Introduction

Polygraph testing has a long and contentious history in the arena of social science. Perhaps nowhere has there been more controversy than that which surrounds screening uses of the polygraph test. Advocates emphasize the utility of information gleaned through polygraph testing while those opposed question the validity of the field results and the generalizability of analog studies. A comprehensive review of diagnostic and screening applications of polygraphy by the esteemed National Academies of Science (NRC, 2003) reported results that could be

argued as supportive of the opinions held by both proponents and opponents of polygraph testing. Opponents also question the construct validity and ethics surrounding how consumers use polygraph results, and suggest that polygraph results may be unrelated to desired outcomes.

The application of polygraph in a public safety pre-employment screening is complex from the outset, and is best understood by starting with brief review of the putative psycho-physiological concepts that underlie polygraph techniques, procedures, and test interpretation. These issues should

¹American Association of Police Polygraphists

²Boise State University

³APA Past President

⁴Private Practice

⁵Institute for Personality and Ability Testing, Inc.

Authors' Note: The authors are grateful to Dr. Michael G. Aamodt, Dr. Stuart Senter, and Ben Blalock, for their thoughtful reviews and comments to earlier drafts of this paper. The views expressed in this article are solely those of the authors, and do not necessarily represent those of the Department of Defense, the American Association of Police Polygraphists, or the Institute for Personality and Ability Testing, Inc. Questions and comments are welcome at polygraphmark@gmail.com.

This article was originally published in the *Journal of Police & Criminal Psychology* (2009), vol. 24, issue 2, pp 69-86, and is reprinted here with the kind permission of that journal. Minor editing of the original work has been done to bring the article into conformity with the style and format of the journal *Polygraph*.

be understood in the more general context of the inherent complications that are common to all screening tests. Because it is unrealistic to expect perfection from any test, whether medical, psychological, or investigative, informed consumers of polygraph test results should become familiar with common testing concepts including sensitivity, specificity, hit-rates, miss-rates, false-positives, false-negatives, and base-rate influences along with the ways that these may affect polygraph screening outcomes. We suggest that validity and reliability of current polygraph field practices may be improved through increased emphasis on the selection of polygraph examination targets for which there is evidence of their actuarial contribution to desired outcomes. We further suggest polygraph results and information may be most effectively employed in the context of a “whole person” approach to evaluating law enforcement applicants. This approach may be best guided by the efforts of police or industrial psychologists whose training in psychodiagnostics and empirical methods will allow them to effectively navigate potentially positive and negative aspects which the polygraph offers to the police personnel selection process.

Our ultimate goal is to suggest the polygraph can be a valuable tool, at the disposal of the police or industrial psychologist, to help them make better recommendations to law enforcement hiring officials faced with the difficult puzzle of determining just who to screen-in or screen-out of the law enforcement selection process.

Background and history of polygraph screening programs

Today, pre-employment screening polygraph examinations of police applicants are widespread in the US and elsewhere, and are intended as an aid in the selection of suitable applicants. Unlike diagnostic tests, which are used for criminal investigation polygraphs, screening examinations are conducted in the absence of any known incident or allegation. Screening polygraphs and screening tests in general, are often constructed to investigate, in a cost effective and expedient manner, the applicant’s history of involvement in a range of possible activities of concern to hiring officials.

Polygraph screening was used to screen employees as early as the 1930s when Leonard Keeler signed an agreement with the insurance firm Lloyds of London to periodically test bank employees for embezzlement (Alder, 2007). By the 1940s, polygraph screening tests were conducted on German prisoners of war for potential post-war law enforcement positions (Linehan, 1978). One of the earliest large-scale testing programs was that of the Manhattan District of the Corps of Engineers which began vetting potential employees for the Oak Ridge nuclear weapons facility in 1946 (Linehan, 1990). This testing program was considered by some to be successful in that it contributed to the return of many previously stolen tools and supplies and elicited admissions of serious work related transgressions such as unreported spills of radioactive materials. The Atomic Energy Commission, however, discontinued the polygraph screening program in April, 1953, in part because the program was seen as providing only marginally increased security (Krapohl, 2002). Polygraph screening gained popularity in the United States private sector during the 1970s and 1980s. As many as 2 million Americans a year were being tested, mostly in the private sector, by the 1980s (Alder, 2007).

The US Congress enacted the 1988 Employee Polygraph Protection Act (EPPA) to curtail among other things, abuses reported as a result of the widespread use of polygraph. Problems observed prior to EPPA included poorly standardized and unregulated field practices, and inadequately standardized training for field practitioners, and included cost-cutting and other competitive marketing efforts that led to the proliferation of “chart-rolling” practices which included the conduct of numerous short and unreliable examinations. Foremost among those problems was the selection of examination targets with unproven contribution to the desired outcomes of employee training success and employee integrity. Despite the restriction imposed by EPPA, there are remaining provisions that allow for government and public safety pre-employment polygraph screening (Krapohl, 2002), in addition to potential screening for employees in pharmaceutical and nuclear energy industries.

Meesig and Horvath (1995), in conjunction with the American Polygraph Association (APA), conducted a survey to determine the use of pre-employment polygraph testing in 626 law enforcement agencies throughout the United States. The mean force size of the agencies surveyed was 447 officers, serving an average population of 522,000 citizens. The survey found that approximately 62% of the respondent agencies utilized the polygraph as part of their hiring process. The respondent agencies reported that they rejected approximately a quarter of their applicants as a result of information produced through polygraph testing that had not been uncovered with their other screening processes.

The Meesig and Horvath (1995) survey revealed that previous illegal drug use was the main content of the information gathered as a result of polygraph, but criminal activities were also disclosed. Respondent agencies reported that the polygraph screening uncovered information indicating involvement by some applicants in unsolved homicides (9%), perpetration of rape by applicants (34%), and commission of armed robberies (38%). The majority of these agencies felt polygraph testing was as useful as (or better than) other forms of vetting, including background investigation, written psychological tests, psychological or psychiatric interviews, personal interviews, and interviews by a selection board.

The U.S. government is arguably the largest user of the polygraph (Krapohl, 2002; NRC, 2003). Government polygraph screening programs have steadily increased over time, and there are presently in excess of 20 federal polygraph programs dedicated to screening applicants, employees, and contractors for access to sensitive information. According to Barland (1999), 69 countries around the world have known polygraph capabilities and that number is almost certainly larger today. Polygraph screening programs are in place in both private and public sectors in the United States, Mexico, Israel, Japan, South Africa, Bulgaria, Russia, and Canada (Krapohl, 2002).

Legal history surrounding polygraph testing in the United States court systems

Legal admissibility of polygraph test results in the U.S. court systems has a long and colorful past. Perhaps no other evidentiary offering has been scrutinized to a greater degree than polygraph test results and admission of polygraph test results into legal proceedings is rare (Daniels, 2002). Concerns include whether the polygraph evidence would overwhelm, confuse or supplant the trier of the fact. Additionally, issues of validity and reliability of polygraph testing in general are bound to be raised. The popularity and allure of polygraph testing has left no dearth of studies from which one may report results. For example, the National Research Council (2003) reported results from 50 laboratory studies which met their criterion for inclusion in quantitative analysis and that alone included 3,099 polygraph examinations. A number of published studies were excluded because they did not meet their criteria for inclusion (NRC, 2003).

American polygraph law was the impetus for the "General acceptance in the scientific community" test which has been referred to as the "Frye test" in honor of the case that set the precedent, *Frye v. United States* (Daniels, 2002). Defendant Frye was convicted of murdering a prominent Washington, D.C. physician in 1920 (Krapohl & Stern, 2003b). Frye appealed his conviction based on the trial court's refusal to admit the results of a discontinuous systolic blood pressure "deception test" administered to Frye by Dr. William Marston (Daniels, 2002; Krapohl & Stern, 2003b). The deception test was purported to be able to determine veracity based on periodic sampling of the examinee's systolic blood pressure during questioning about the crime event. This case occurred at a time in history when judges and courts were being presented with offerings of new scientific based evidence, but often without the benefit of testimony on acceptance from the scientific community. The court of appeals upheld the trial court ruling to not allow Dr. Marston's testimony regarding the deception test he administered Frye and in doing so established a precedent for novel scientific evidence that endured for the next 70 years (Daniels, 2002).

This “Frye test” required a scientific test to have gained the general acceptance of the scientific community in the particular field from which it belongs.

Several more recent court opinions appear to allow some opportunity for polygraph admissibility. In 1989, the federal appeals court in the 11th circuit, opined “a per se disallowing of polygraph evidence is no longer warranted” in the case of *United States v. Piccinonna*. This decision still stands as it has not been overruled, but it has not been followed by any other federal jurisdiction (Daniels, 2002). The other recent case potentially affecting polygraph admissibility is 1993 United States Supreme Court decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (*Daubert*). While not addressing polygraph specifically, *Daubert* addressed the “Frye test” and found it to be too restrictive. The United States Supreme Court essentially stated in the *Daubert* opinion that admissibility revolves around a number of factors which when taken as a whole allow lower courts some flexibility in admitting evidence that results from emerging scientific developments.

These factors include:

1. is the theory being offered capable of being tested;
2. has the error rate for the application of the technique been established;
3. has the application of the technique been subjected to peer review and publication;
4. is there a known level of acceptance of the offered theory by the scientific community most relevant to the particular theory;
5. are there established standards to determine the correct and acceptable application of the technique.

The overwhelming trend has been to exclude polygraph evidence from American courtrooms by applying a stringent *Daubert* interpretation and it seems unlikely that any significant change to admissibility will occur in the near future (Daniels, 2002).

General information about polygraph testing

Though they operate in different domains, screening polygraph examinations

share many of the testing principles with diagnostic polygraph exams used in criminal investigations. Commonalities include the basic principles of question formulation, testing protocol, and instrumentation. There are, however, important differences between screening and criminal investigation or diagnostic polygraphs. Criminal investigative polygraph examinations focus on an examinee’s involvement in a known event or known allegation, whereas screening examinations test for credibility about involvement in specified patterns or categories of behavior, over sometimes lengthy time periods, which are empirically correlated with increased risk for an undesired future outcome. For example, a question from a diagnostic or investigative polygraph test might be, “Did you rob the 1st National Bank last November 4th?” whereas a typical screening question might be worded as “Did you ever commit a serious crime?” The time period for screening exams is necessarily broader, and instead of referring to a known allegation or known incident pertaining to a specific date or period of time, screening exams may refer to the examinee’s entire lifetime or entire adult lifetime. Screening test questions may also be limited to a recent period of time that will improve the signal value and actuarial utility of the target information. For example: “During the last five years, have you had any involvement with illegal drugs?”

Polygraph test questions should provide interpretable and useful information to the consumer of the test result, regardless of whether the examinee passes or fails the test. For test questions to meet this requirement, it is necessary that all target questions meet certain commonly accepted criteria, including that the question describe the examinee’s possible involvement in a single behavior or single pattern of behavior, can be easily answered ‘yes’ or ‘no,’ does not include vague or necessary legal or clinical jargon, is free of references to motivation or intent, and does not presuppose guilt or involvement on the part of the examinee. Behaviors referred to in polygraph test questions should be supported by a operational definition that is commonly understood between the examinee, examiner, and the referring professional. Operational definitions provide descriptive information

about what one would be observed doing if one were to engage in a behavior. This has important implications for the validity and utility of the polygraph testing. For example, one could be expected to know with certainty, whether or not one had robbed a bank. The same degree of assurance may not be attributable to a question requiring the examinee to search their memory for whether they committed a serious crime. It would be a problem, for example, if an applicant does not understand the hiring agency's operational definition of what constitutes a serious crime versus a non-serious crime. An examinee in such an ambiguous situation, without an adequate operational definition, would be faced with the task of answering a question about involvement in serious crimes while at the same time trying to identify what separates a serious from non-serious crime.

As with all tests that render dichotomous outcomes, there are two correct outcomes and two types of errors that can occur with the polygraph. A positive result signifies the examinee's involvement in the behavior or behavioral category described by the relevant question. Similarly, a negative result suggests the examinee was not involved in the behavior or behavioral category of concern. When a truthful examinee is judged to be deceptive by a polygraph examiner the error is called a false positive error, or more simply a false-positive. Conversely, when a deceptive examinee is judged to be truthful it is a false negative error or false-negative. Along those lines, a true-positive result is one in which a deceptive examinee is correctly identified as deceptive and a true-negative would describe a truthful examinee judged to be telling the truth. The desired attributes of a polygraph test are identical to the goals of other forms of testing. It is a requirement of all effective tests that they provide high enough sensitivity to reliably notice the issues of concern, thereby avoiding false-negative errors. Another desired characteristic of effective tests is that of providing high specificity to the issues of concern, ensuring unrelated factors will not cause false-positive errors. Unfortunately, there is no such thing as a perfect test that can perfectly accomplish both objectives of sensitivity and specificity. In every form of testing, there is always a trade-off or compromise among these objectives. Test developers have learned to make strategic

compromises in the design of test protocols, in order to differentially prioritize these objectives in diagnostic and screening test contexts. It is important for administrators and consumers of polygraph results to remain aware that a small portion of errors of some kind can always be anticipated from any test or procedure.

Use of decision theoretic approaches in polygraph practice has a short history, and until recently, there were no serious attempts among practitioners to develop a body of best practices for screening examinations. This is an unfortunate circumstance that has led to serious problems for the profession as a whole. The lack of practice standards, and other inadequacies, may have contributed to the passage of EPPA in 1988 which severely restricted polygraph usage outside of government agencies, as discussed earlier. Moreover, research has confirmed that potential inadequacies, such as low sensitivity and/or specificity, existed in polygraph screening methods employed at the time, even in the much better controlled environment of the U. S. Government (Barland, 1981; Barland, Honts, & Barger 1989; Honts, 1992).

One current challenge in the field is a lack of standardization in test administration across the profession. In recent years, a series of articles in professional journals and other venues urged the polygraph profession toward more data-driven field practices (Harris, Horner, & McQuarrie, 2000; Kircher, Kristjansson, Gardner, Webb, 2005; Krapohl, 2006; Raskin & Honts, 2002). As an example of the trend away from values-based and idiosyncratic field practices, the largest polygraph professional organization is currently developing model policies, including those for polygraph screening of police applicants and other specialties (American Polygraph Association, 2008). The American Society for Testing and Materials (ASTM, 2008) has promulgated standards for a variety of polygraph tests and settings. Though not a complete solution to the problems of practice standardization, the efforts of these organizations represent a significant step toward embracing standardization principles found successful in other fields such as medicine and psychological assessment.

Diagnostic tests are intended to help formulate a basis for necessary action, and should provide sufficient specificity to the issue of concern to accurately identifying persons not involved in the issue under investigation, thereby reducing an investigator's burden of suspicion. In actual field practice, decision schemes for diagnostic polygraph tests are typically risk-averse. That is, they are deliberately set to ensure that a guilty suspect remains on the investigative radar.

Screening tests should also be designed to be risk-averse, and strive to reduce the likelihood that a problem goes undetected. When the consequences of an erroneous judgment (e.g. that an unsuitable candidate is hired into a police role) present potentially catastrophic ramifications to an agency or community, decision thresholds should be set to maximize detection sensitivity to potential problems.

Krapohl (2002) and Krapohl and Stern (2003a) discussed the differences between screening and diagnostic polygraphs, including the use of a successive-hurdles model (Meehl & Rosen, 1955) when seeking to mitigate decision errors and maximize the effectiveness of polygraph testing programs. Diagnostic and screening tests are used in many fields, and when thoughtfully combined in the screening domain, these two distinct testing approaches may offer unique advantages to both decision makers and consumers of test results. Screening methods are generally intended to be a cost effective, though imperfect means of sorting individuals into tentative categories. Although diagnostic methods may have substantially more classification power than screening methods, they also tend to be more resource-intensive, and are therefore more wisely reserved for only those individuals who produce positive results on the screening tests. Screening tests are therefore useful only when they provide adequate sensitivity to the issue or issues of concern. In an effort to maximize the sensitivity levels of screening exams, test developers adopt decision thresholds that provide adequate sensitivity to reliably identify the presence of the issue or issues of concern. A process model, that includes the use of diagnostic testing only after a positive screening result, reaps benefits to the end-

user by improving overall decision accuracy over mere screening, while mitigating costs over the use of multiple diagnostic tests. The effect of ensuring that every possible unsuitable candidate is identified and eliminated from the pool of eligible applicants will inevitably result in the elimination of some suitable candidates. A comprehensive program would respond to all positive test results (those signaling a significant response) with additional or follow-up investigative procedures in response to this known and expected reduction of test-specificity in screening situations (NRC, 2003). Ideally, such follow-up investigative procedure would employ methods that offer better specificity than the initial screening exam, in an effort to reduce the incidence and impact of false positive results. These investigative responses may include follow-up polygraph testing with more specific procedures or additional background investigation efforts aimed at clarifying the issue of concern. This method may at first seem administratively cumbersome but the advantages become apparent when one considers the costs and expense of intensive investigative procedures, compared with the expediency that screening methods provide.

Polygraph Test Techniques

Polygraph techniques can be divided into two major categories, knowledge-based tests, also called recognition tests, and deception based tests. The knowledge-based tests attempt to determine if the examinee has knowledge only available to persons directly involved in an incident of concern. These tests are commonly known as Guilty Knowledge Tests, or more correctly, as Concealed Information Tests (CIT). In a CIT used for a murder case, the polygraph examiner might assess whether or not the examinee reacts physiologically to the murder weapon as compared to a series of possible weapons which investigators are certain were not used in the crime. Because this approach depends upon the existence of a known crime or incident facts that remain unknown to the innocent suspect, the CIT testing paradigm is not suited for use as a screening test concerning unknown incidents and multiple issues, and will not be discussed further here.

Polygraph screening programs generally rely on deception-based methods. These methods ask directly about the matter to be assessed, are capable of addressing multiple behavioral issues of concern and do not depend on the existence of a known incident or known allegation. These features mean that these tests are suited for the screening environment, they are intended to assess an examinee's credibility regarding involvement in behaviors of concern, or conformity to personnel selection standards. While the circumstances and case facts of a criminal investigation drive the selection of polygraph questions in a very straightforward manner for diagnostic polygraph examinations, the issues for police pre-employment screening polygraphs are usually driven by department policies. Unfortunately, these policies are often more tied to preferences of the leadership than to empirically derived predictors, a factor which almost certainly limits the value of the polygraph screening programs. Ideally, personnel hiring policies would be informed by actuarial data concerning successful training and job performance outcomes. An actuarially based polygraph screening program should deliver information of better predictive value than is generally found among current police polygraph screening programs (Aamodt, 2004).

There are two broad categories of deception tests, the historically older Relevant-Irrelevant and the Comparison Question Tests (CQT). The Relevant-Irrelevant (RI) test involves asking direct questions, known as relevant questions, about the matters to be assessed (e.g., Did you ever commit a serious crime?). The RI test also contains several simple, known-truth questions that are usually answered truthfully (e.g., Are the lights on in this room?) known as irrelevant questions. The questions are repeated several times while the examinee's physiology is monitored. The rationale of the RI test assumes that deceptive individuals will respond with consistent and significant physiological response to those questions to which they are deceptive, whereas the truthful examinee will not show such responses. In general, the evaluation of RI polygraph exams calls for the examiner to make an interpretation of what the terms consistent and significant mean while evaluating the test

data. This has the potential to degrade any inter-rater agreement in the evaluation by introducing subjectivity.

Raskin and Honts (2002) concluded that the rationale of the RI technique is naïve, and that the approach does not presently satisfy the basic requirements of a psychological test and should not be used in forensic/investigative settings. There is, however, some evidence that shows the RI approach to screening may have validity (Correa & Adams, 1981; Honts & Amato, 2007; Krapohl, Senter, & Stern, 2005). In the screening context, the RI test may be suitable as an early screening tool in which the objective is to investigate multiple relevant topics. More data are needed to make strong statements about the validity of the RI test in the screening setting. However, new approaches to computer-based data analysis (Kircher, Woltz, Bell, & Bernhardt, 1998) and test automation (Honts & Amato, 2007) may well raise the level of validity for the RI test sufficiently to make it a viable choice for screening applications.

The second family of deception tests, the CQT, uses relevant and irrelevant questions similar to those used in the Relevant-Irrelevant test, but also includes a third type of question, the comparison question. Comparison questions are designed to evoke responses from innocent individuals and to provide the innocent person a place to focus one's emotionality and attention. In the CQT an interaction is expected between the physiological responses to question type (relevant and comparison) and guilt status. Guilty examinees are expected to produce larger physiological responses to relevant questions than to comparison questions. Innocent examinees are expected to show the opposite pattern. There are many field testing formats that fall within the CQT category, most of which are named after the agency or surname of their creator. In terms of administration, differences between CQT formats are trivial, mostly surrounding question ordering. The more important non-trivial differences between CQT formats applied in field practices concern the analyses, and we return to those issues below. Readers interested in the differences between the CQT variants are referred to Raskin and Honts (2002).

Within the CQT category, an additional distinction is that there are two approaches to the preparation and presentation of the comparison questions: the probable lie comparison (PLC) and the directed lie comparison (DLC). With PLCs, the examinee is maneuvered into denying transgressions generally related to those addressed by the relevant questions. For example, if the relevant questions address the theft of a ring, a PLC question might be, "Have you ever stolen anything from a place you worked?" The examiner discourages the examinee from admitting such transgressions by implying that the person who stole the ring would have a similar pattern of thievery at other places, including their job or the theft of the item in question. This social dynamic pressures the examinee to lie to the PLC question to avoid presenting the appearance of the kind of person who stole the ring. The rationale is that the innocent examinee will be more focused on, and hence more physiologically aroused, when lying to the PLC questions than when being truthful to the relevant questions (Offe & Offe, 2007, Raskin & Honts, 2002). It is further theorized that the actual thief will be more aroused physiologically by the relevant questions than by the PLC questions to which he or she is also lying. There is substantial body of laboratory and field research in the forensic setting that supports this rationale and the validity of the PLC version of the CQT (see the reviews by Honts, 2004; Raskin & Honts, 2002).

The DLC approach is a simpler and less manipulative approach in which the examiner instructs the examinee to lie to questions similar in form to PLC questions (Honts & Raskin, 1988; Raskin & Honts, 2002). The examinee is told that it is important for the examiner to observe appropriate physiological responses from the examinee whenever lying to the DLCs during the test, otherwise the test will be inconclusive (Raskin & Honts, 2002). DLC questions offer advantages over PLC questions, such as standardization of test questions and avoidance of manipulative processes during the examination. PLCs may impede the building of rapport during the examination, whereas this is less likely to occur with DLCs. Finally, examinees with prior polygraph experience or those who have researched polygraph techniques may be aware of the

PLC procedures which would serve to reduce any potential face validity imparted on the PLCs to the examinee by the examiner. This is not a problem with the DLC as screening use was one the situations for which the format was originally developed (Menges, 2004). The rationale underlying the DLC approach is similar to that of the PLC and the same interaction of question type and guilt is expected. Although there is less scientific research on the DLC test, the existing data suggest it is of equivalent or higher validity as compared to the PLC test. Indeed, the Test for Espionage and Sabotage (Research Division Staff, 1997; 1998; Reed 1994), a DLC technique, was shown to outperform PLC screening techniques, and is among the small number of techniques reported by (Horowitz, Kircher, Honts, & Raskin, 1997; Krapohl, 2006) to meet a list of specific criteria for recognition as validated methods.

Interpretation of Polygraph Tests

There are important differences in field practices in the interpretation of polygraph data. Ideally, it would seem polygraph examiners should use analysis procedures based on scientific study and reasoning. This would afford a valid basis for their decision making. Interpretation of polygraph data from screening tests is sometimes done with a global analysis, in which the examiner visually inspects the test data for physiological reactions or changes in reaction trends that occur in response to repeated presentations of a particular question. The global analysis approach is limited to visual pattern-recognition approaches and the examiner's clinical impressions regarding whether the examinee has been truthful or deceptive to the relevant topics. Such analyses are known to be less valid than more systematic objective approaches (Kircher, Horowitz & Raskin, 1988; Raskin, Barland & Podlesny, 1978). Research has supported the scientific assumption that clinical global analyses of polygraph data should be inferior to more actuarial approaches (Raskin & Honts, 2002).

The most common field approach to evaluating polygraph data is known as numerical scoring. In numerical scoring, the examiner performs a series of magnitude comparisons and assigns scores based upon putatively validated criteria. Decision making

with numerical scoring of CQT polygraph data involves the simple cumulation of scores that are then evaluated against decision cutscores. Cutscores for determining truthfulness or deception have been the focus of numerous investigations regarding their empirical validity with diagnostic polygraph tests (Capps and Ansley, 1992; Harwell, 2000; Krapohl, 1998; Krapohl & Cushman, 2006; Senter & Dollins, 2003; Van Herk, 1991). However, little effort has been made to study the effect of adjusting cutscores in screening tests.

There are also commercial computer-based statistical approaches to the analysis of CQT polygraph data. Those approaches based upon multivariate classifiers usually return the probability that the current test examinee's data belongs to known standardized distributions of deceptive or truthful cases. Unfortunately, all of the commercially available computer scoring algorithms were standardized using data from forensic, rather than screening settings. Thus, the generalizability of these algorithms to the screening setting is unknown. Nelson, Handler and Krapohl (2008) have recently developed an open-source computer scoring algorithm that includes a statistical and decision model designed to manage the complications inherent to mixed-issue screening polygraphs. Additional research is needed to demonstrate that end-users can validly infer probability of truthfulness or deception from the result of a screening polygraph test using computer models designed to handle that type of data. Despite the availability of computer scoring systems and their superior reliability compared with humans when completing complex inferential calculations (Kircher & Raskin, 2002), field polygraph examiners have continued to rely primarily on hand-scoring systems, point totals, and cut scores to formulate their opinions.

Polygraph examiners are trained to report the results of criminal investigation polygraphs using the terms "deception indicated," or "no deception indicated," and often use the terms "significant reactions" and "no significant reactions" for screening exams (ASTM, 2002; Research Division Staff, 2006). The terms deception indicated and significant reactions are synonymous with a conclusion that the examinee has been deceptive, while

the terms no deception indicated and no significant reactions are synonymous with a conclusion that the examinee has been truthful to the test questions. Examiner decisions are the result of numerical scores and established numerical decision thresholds.

Polygraph evaluation approaches allow for an opinion of inconclusive, and examiners who are familiar with federal training protocols may use the term no-opinion in a synonymous manner (Research Division Staff, 2006). Inconclusive and no-opinion results occur when insufficient definitive information exists to make a decision of truthful or deceptive. One view is that an inconclusive result represents an error inasmuch as the polygraph failed to correctly make a classification. Another view is that an inconclusive outcome is not an error but indicates a reduction in utility. That is, although an inconclusive result cannot be used to support a professional opinion, no harm is done by such classification. From a practical perspective, an inconclusive result simply means that something has gone wrong (e.g., examinee fatigue, confusion, deception and even countermeasures), and that the test result cannot support a professional opinion. In programs faced with the need to inquire about very serious issues, such as crimes against the country, it is important that an inconclusive outcome should not become a terminal event in the selection process. Honts (2005) found that inconclusive results occur with greater frequency among truthful examinees than deceptive examinees. Nelson, Krapohl and Handler (2008) described that an asymmetrical alpha scheme, using $\alpha = .1$ for truthful decision and $\alpha = .05$ for deceptive decisions resulted in reduced inconclusive results with only minimal increase in false-negative errors. Because an inconclusive result cannot deliver a result that is much more predictive than simply using the base rate as a guide, re-testing the examinee should be viewed as the best-practice. We caution that neglecting to retest an examinee who produced an inconclusive result, in programs where policies bar an individual from further consideration for employment due to unresolved polygraph test results, will most likely be perceived as a de facto test error.

The debate about the decision theoretic meaning of inconclusive outcomes is beyond the scope of this paper. In short, our view is that inconclusive results do not contribute to false-positive or false-negative outcomes, and contribute neither to sensitivity to deception nor specificity to truthfulness. It is for this reason we will set aside inconclusive outcomes and consider only conclusive decisions.

Unique Characteristics of the Polygraph Screening Context

Forensic-investigative polygraph examinations focus all of the relevant questions on a single known incident or known allegation, whereas screening polygraph tests involve the simultaneous investigation of multiple behavioral or suitability topics in the absence of any known incident or allegation. It is relatively easy to understand that the presence or absence of reactions to any or all of the test questions of an investigative polygraph would signal involvement or non-involvement in a single known incident. Because screening polygraphs, involve multiple distinct issues, they include the possibility that an examinee could be deceptive to one or more relevant issues while simultaneously being truthful to others.

The limited research on multiple issue screening polygraphs suggests that polygraph test results can detect when an examinee is being deceptive. However, the research does not support the idea that polygraph tests can pinpoint the exact issue about which an examinee is lying (Barland et al., 1989; NRC, 2003; Podlesny & Truslow, 1993). Part of the problem is psychological: an examinee will not always react to every question to which they are lying, but may react only to the questions that the examinee perceives to be most important. As mentioned, all polygraph examinations, even mixed issue screening exams, are a form of omnibus test. This means that decision rules associated with numerical scoring require that for an examinee to be considered truthful during the polygraph examination, they must produce truthful scores to all test questions, while a decision of deception requires failure to only one of the questions (Research Division Staff, 2006). As such, mixed-issue polygraphs

represent a forum of testing in which the truthful results are subject to all-or-nothing interpretation rules and this aspect of polygraph testing is not well appreciated outside of the polygraph profession. It is helpful for police psychologists and personnel administrators to understand these limits of polygraph testing so not to misinterpret what test results imply.

Handler, Nelson and Blalock (2008), in an attempt to address some of the perceived inadequacies of CQT screening techniques (i.e., lack of standardization, numerous target issues, biased scoring rules), introduced the Directed Lie Screening Test (DLST) for law-enforcement applicant screening, post-conviction sex offender testing. This procedure is an adaptation of a currently used and researched screening test (TES) described earlier in this paper. Advantages of this approach included improved standardization in test administration, a reduction of relevant questions to two primary issues, and decision policies requiring the examinee be regarded as responding significantly to the examination as a whole, rather than to individual questions. Test protocols allow for the inclusion of additional relevant questions in a separate series of questions. Blalock, Nelson and Handler (2009) describe the automated use of the DLST in field polygraph testing of post convicted sex offenders and public safety pre-employment screening.

Use of the Polygraph in the Employment Selection Process

It can be said that there are two general testing approaches with selection tools: screening-in and screening-out. Screening-in refers to those methods by which employers test applicants for the competencies needed to perform well in an organization. Screening-in assessments include tests of knowledge, skill, and ability. Screening-out, in contrast, is the process of identifying vulnerabilities that would make a candidate a risk to the potential employer. Screening-out assessments include background investigations, credit checks, drug screens, medical exams, and tests of psychopathology. It is important to note that in the screening-out process, the presence of a problem suggests a potential risk, but the absence of a problem does not imply a high

level of future performance. Polygraph testing appears to function better as a screening-out tool, where it is used to uncover behaviors that are incompatible with the integrity required of police officers.

Decision-support is the most obvious goal of a police pre-employment polygraph screening program. Any policy or field practice in which a hiring decision is based solely on the results of a polygraph test, or any single test, would be difficult to justify. A more prudent course is to integrate polygraph results with other sources of information on police candidates, weight them properly among those sources, and base hiring decisions on the totality of the information. This is sometimes referred to as a “whole-person” approach or concept. A whole-person review would take into consideration all aspects of the applicant’s benefits which are then weighed against any risks in an effort to determine suitability. Such an integrative approach could help avoid the problem of over-reliance or under-reliance on the polygraph, or any other single source available to hiring officials.

In addition to polygraph test results, the screening polygraph context facilitates self report of information from candidates that may be of interest to risk evaluators, risk managers and hiring administrators. Many departments find that the polygraph examination process develops more significant information than all other sources combined and some departments may consider the polygraph its most important screening tool for this reason (Krapohl, 2002; Messig & Horvath, 1995).

One final benefit of a police pre-employment polygraph screening program is that it may deter less-suitable candidates from applying for positions in police work or public service. Consequently, those candidates who do apply may be more likely to meet the hiring standards than if the polygraph deterrent were not in place.

Programs with properly integrated polygraph policies will be able to increase the likelihood that police agencies hire suitable candidates, while simultaneously decreasing the likelihood that unsuitable applicants are hired. Much has been made of the limits of

the polygraph and how it may lead to the disqualification of suitable applicants (Lykken, 1998) but consider the following thought experiment where polygraph testing is removed from the screening process. First, it is important to recognize that a large body of research indicates that human detection of deception is barely above chance levels (54%) and that law enforcement professionals are no better at detecting deception than the public in general (Aamodt & Custer, 2006; Vrij, 2008). Thus, it is an unfortunate consequence that applicants with and without concealed disqualifying behavior in their backgrounds have about the same probability of being selected or rejected by a department, assuming the disqualifying behavior is not discoverable by other means. Let us now include the polygraph in the equation. A polygraph screening program with sensitivity and specificity rates above chance will improve the prospects for the truthful candidate by reducing the number of unqualified candidates.

Polygraph and the Americans with Disabilities Act

Although polygraph examinations are used in both the pre-conditional and post-conditional offer stages in law enforcement hiring, the Americans with Disabilities Act (ADA) limits the scope of pre-offer medical inquiry. Polygraph examiners often inquire as to the examinee’s psychological and medical history when attempting to determine suitability during the pre-test phase of the polygraph exam and before test data is collected. An important consideration to both administrators and polygraph examiners involves whether to conduct a pre-employment polygraph examination prior to or after a conditional offer of employment has been given to the applicant. Some of the lines of questioning historically taken during pre-employment screening may not be in compliance with ADA rules (historic alcohol or drug use). Agencies that perform pre-offer polygraph testing would be wise to consult with their hiring counsel in order to ensure their test coverage is in compliance with ADA rules.

Estimated Accuracy of Polygraph Screening

Several studies have attempted to estimate the accuracy of the polygraph for screening purposes. The National Research Council (NRC) addressed the difficulty in making assumptions in accuracy within the context of screening, commenting on the lack of pre-employment polygraph research. The NRC (2003) reported results from four studies (Barland, Honts, and Barger, 1989; Research Division Staff, 1997; 1998; Reed, 1994) they felt met their minimum quality standards. The relevant questions in these studies addressed specific programmed acts on the part of some examinees that were similar to those that might be asked in a screening polygraph test in a national security setting. For example, some examinees were assigned to the deceptive condition wherein they performed simulated acts of espionage or sabotage. Barland, Honts and Barger (1989) involved three separate experiments. In their first experiment, the false negative rate was 66%. In other words, about two-thirds of the programmed guilty examinees passed their polygraph test. Their second study attempted to determine if several tests with single issues was superior to one test of multiple issues. An additional goal of this study then was to determine whether the polygraph test was as good at identifying the deceptive issue(s) with examinees as it was at detecting deceptiveness among (between) examinees. In other words, in a multiple-issue screening situation, is the polygraph as effective at identifying lies as it is in identifying liars? These studies suggest that that within-examinee detection was poorer than between-examinee detection. The examiners were effective at identifying those who were lying but did not as effectively identify the correct questions to which they lied. This is a potential problem since conveying results to the hiring agency and/or psychologist might not yield specific enough information to be included in the entire battery leading toward suitability. Their third study retested a number of individuals from the first study and NRC noted concerns of confounding.

The NRC reported they calculated an accuracy index (A) for only one screening study (Reed, 1994). NRC found for that study, the A was 0.90, corresponding to a sensitivity

(true positives) of approximately 85% and a specificity (true negatives) of approximately 78% (NRC, 2003). The published inconclusive rate for the technique used in this study, the TES, is 2% (Krapohl, 2006). Consumers of polygraph should pay attention to inconclusive rates as they speak to the utility of application of a particular test. A test with a large inconclusive rate may not be amenable to a screening environment, even if it has tremendous accuracy. One other screening study (Honts & Amato, 1999) dealt with deception on pre-employment screening tests. Honts and Amato (1999) looked at automated polygraph testing using the relevant-irrelevant (R&I) testing format. They reported accuracy rates of 77.5% for the automated condition of testing and 65% for the human condition of testing for dichotomous decisions of truthfulness or deception.

The Base Rate Problem

A complication in interpreting screening test results, or any diagnostic screening for that matter, is the base rate problem. The term base rate refers to the frequency with which the target of a diagnostic test appears in the population being tested. In the context of police applicant screening, the base rates of the various targets (relevant questions asked) of the screening polygraph test may vary widely. For example, one target of the screening polygraph may be to determine if there are undiscovered/undisclosed felonies. The rate at which these occur in traditionally vetted police applicants referred for polygraph screening is in all likelihood low. However, a target such as undisclosed use of illegal drugs may have a higher base rate.

While the base rate does not directly affect the accuracy of the polygraph test, it does change the confidence that the hiring official can have in the polygraph test outcome. Consider, for example, the following situation adapted from Honts (1991): A polygraph test that is 90% accurate with both truthful and deceptive applicants, and a situation where exactly half of the applicants are truthful. Table 1 illustrates the rates of correct and incorrect outcomes in this situation. Here, with the base rates of truth and deception being equal, interpretation of the outcome is clear. Ninety percent of the

significant response outcomes are correct and 90% of the no significant response outcomes are correct. Alternatively, consider a situation where the majority of the applicants may be attempting deception, which might be the case for an organization that has a zero tolerance for any illegal drug use. Here we have the same polygraph test that is 90% accurate for both truthful and deceptive applicants. However, now 90% are attempting deception

and only 10% are truthful. Table 2 illustrates the correct and incorrect outcomes in this situation. In this case 98.8% of those who produce a significant response on the test were in fact deceptive. However, only 50% of those who showed no significant response were truthful, quite a different situation than for equal base rates and one that is clearly of importance to the persons who make hiring decisions.

Table 1. A conditional probability analysis of 500 truthful and 500 deceptive applicants with a polygraph test that is accurate 90% of the time

Outcome	Truthful	Deceptive	Totals
Significant Response	50	450	500
No Significant Response	450	50	500
Totals	500	500	1000

Table 2. A conditional probability analysis of 100 truthful and 900 deceptive applicants with a polygraph test that is accurate 90% of the time

Outcome	Truthful	Deceptive	Totals
Significant Response	10	810	820
No Significant Response	90	90	180
Totals	100	900	1000

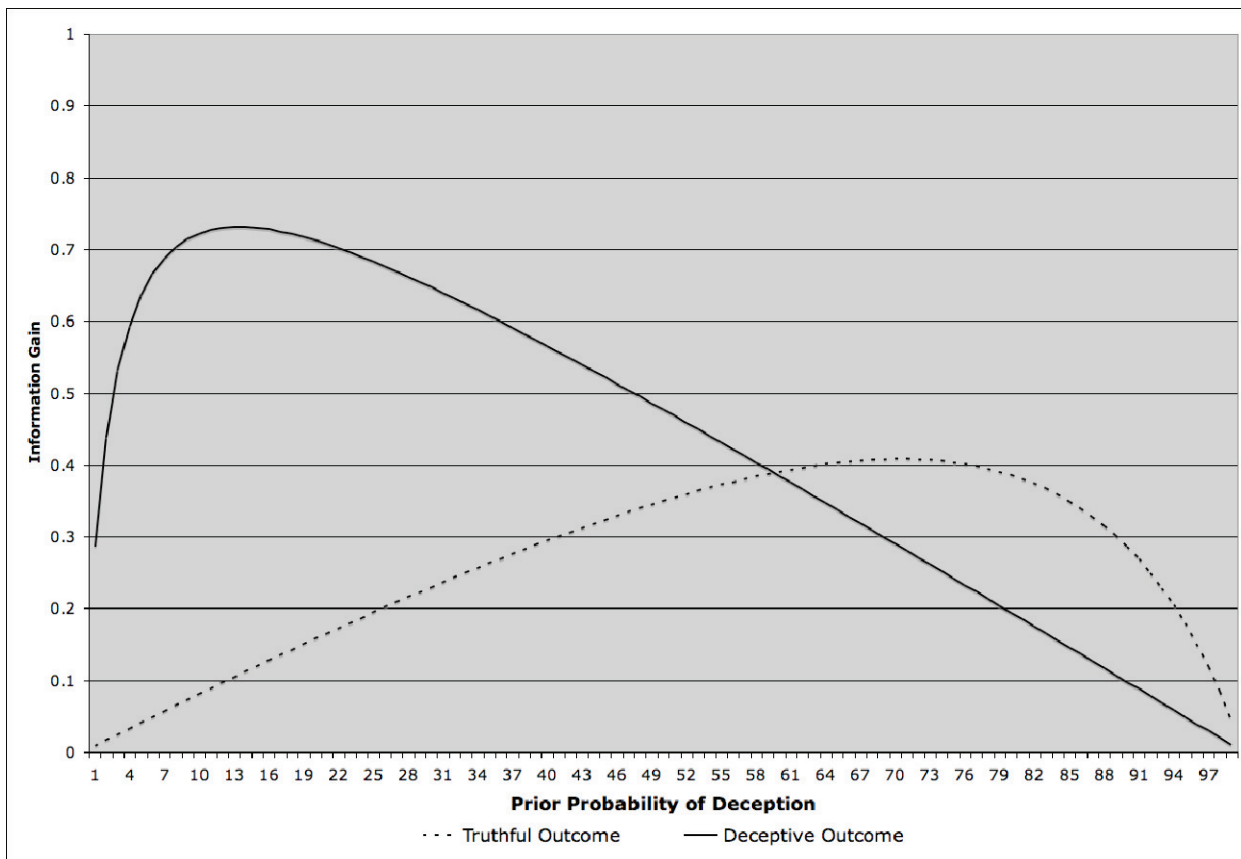
To aid the decision makers under conditions of varying base rate situations, Wells and his colleagues (Wells & Lindsay, 1980; Wells & Olson, 2002) developed a new index, the Information Gain Index (IGI), to illustrate the information gained by using a diagnostic test over chance performance across the range of base rates. Although originally developed for use in the eyewitness

identification setting, Honts (2005) adapted this index for use with polygraph examinations. Figure 1 illustrates a set of Information Gain analyses based on the accuracy figures provided by the United States Department of Defense Polygraph Institute (DoDPI) in a laboratory experiment of the Test for Espionage and Sabotage (TES) (Research Division Staff, 1997; 1998). That study

produced results that were 83% accurate with the deceptive and 96% accurate with the truthful. In interpreting Figure 1 the reader should note that an IGI analysis of unassisted lie detection based upon the massive meta-analysis reported by Vrij (2008) would show a peak information gain over guessing slightly above value of 0.1, that is a 10% improvement over chance (Honts, 2005). With TES data, the IGI curve for deceptive outcomes had a strong positive skew and peaked with a value of 0.72 at a p(deception) of 0.19. This means that with a base rate of guilt of 19%, a deceptive outcome provides a 72% improvement over

chance. For deceptive outcomes the IGI exceeded 0.1 in the range of p(deception) from 0.01 to 0.89 inclusive. The IGI curve for truthful outcomes had a negative skew and peaked with a value of 0.41 at a p(deception) of 0.70. For truthful outcomes the IGI exceeded 0.1 in the range of p(deception) from 0.13 to 0.97, inclusive. These results suggest that even with the modestly accurate TES reported by DoDPI, improvement over unassisted deception detection was achieved over a wide range of base rates for both truthful and deceptive examinees.

Figure 1. An Information Gain Index analysis polygraph screening based upon the DoDPI (2005) study of the TES.



These results provide considerable support for the use of polygraph screening within the likely range of the reasonably common targets of most police applicant screening polygraph examinations. The IGI curves could be a useful guide to hiring decision makers depending upon the

availability of reliable base rate information for their application population. Interested readers should refer to Wells and Olson (2002) for the mathematical detail underlying and the formulae used to calculate the IGI. In practical terms, these IGI curves suggest that the selection polygraph investigation targets with

anticipated moderate base-rates may be most effective. Behaviors with low base-rate or high base-rate occurrence may provide little signal information to assist adjudicators effectively discriminate using the polygraph, between suitable and unsuitable applicants.

Additional Issues

The construction of screening questions that are memorable and relevant to hiring decisions, and those referring to specific types of behavior, will inform the agency about the examinee's suitability or risk. Questions so constructed can improve the ease with which screening polygraph results are interpreted by polygraph examiners and personnel administrators. We suggest that relevant questions should strive to be actuarial in nature, based on known predictors of training and job performance success.

In any employment interview, it seems likely that there will always be more that could be disclosed. What is not clear is if this additional information would add incremental validity to the hiring process. In-depth and all out disclosure may serve the needs of agencies with substantially more applicants than vacancies, or agencies that need to, or can afford to, have highly restrictive hiring tolerances. Such an approach may not serve the needs of smaller agencies with noncompetitive compensation and benefits packages or with fewer applicants. Smaller agencies may be more concerned with the practical differentiation of experimental, recreational, medicinal/instrumental, habitual, or addicted/dependent drug use. Alternatively, they may place more emphasis on the practical matter of how recently someone last used any form of illegal drug. Some agencies have different tolerance levels for past drug use, and may consider hiring someone who stopped using drugs several years ago, even though it may not be realistic to debrief or disclose every single incident in someone's history. It would seem unwise for evaluators or hiring administrators to ever assume they know everything – though this may be more realistic for agencies with highly restrictive hiring tolerances. For agencies with less restrictive hiring tolerances, such as local police agencies that might consider hiring an applicant with some history of recreational

drug involvement as a juvenile or young adult, with some mandatory period of cessation or abstinence, an attempt to know everything would either lend toward pretense or create a condition in which an applicant is externally motivated to exaggerate one's history of involvement with illegal drugs. Polygraph critics will inevitably point out that such exaggeration amounts to making a false-confession.

As described above, we suggest agencies adopt a whole-person approach to hiring rather than draw invariant and arbitrary criteria that determine applicant acceptability. Clearly, the choice to take a whole-person approach and the evaluation of assessments rests on the shoulders of the police or industrial/organizational psychologists and the hiring officials. A criticism of the whole-person approach, as opposed to a rule-based guideline, will be that hiring standards may be perceived as inconsistent and therefore arbitrary. We suggest that the absence of actuarial data that defines for example, the relevance of quantified drug use means that any standardized approach is equally arbitrary.

Just as we caution against the use of polygraph results as the sole basis for police selection decisions, we also caution against the selection of polygraph relevant questions based solely on the value-system or beliefs of any individual examiner or police administrator. As discussed earlier, the questions used during mixed-issue screening polygraphs may encompass a wide spectrum of behaviors in an effort to capture as much relevant information as possible about an examinee's past. There is, however, a limit to the number of questions that can be realistically investigated during a single examination and many modern polygraph techniques attempt to accommodate only up to four relevant questions. One often overlooked aspect is that each additional topic attempted to be covered, is likely to incrementally erode the accuracy of the test. The more issues presented to the individual, the more diluted the person's psychology is likely to become, with results that are difficult to predict within the limits of current knowledge. Clearly this is an area in need of research.

Another accuracy degradation concern surrounds the area of ambiguity of test questions. Polygraph examiners and program administrators must provide adequate operational definition for the test questions so they have a clear and easily interpretable meaning. With questions that lack clarity, physiological reactions measured during a polygraph may have little to do with whether the examinee is lying but may be more the consequence of the mental effort of recall, retrieval or interpretation.

Improving Polygraph Screening through Scientific Question Selection

The selection of polygraph screening questions should provide predictive validity towards risk prediction and job performance (e.g., physical violence, and patterns of undetected criminal activity) for which there may not be a superior alternative method. Unfortunately, police hiring administrators and/or adjudicators may be unaware of the actuarial predictors for successful police work, the limitations of the polygraph, or how to create synergy among the individual pieces of their larger screening process. Consequently, the polygraph's contribution to effective candidate selection may be less than optimal. As a beginning, polygraph programs would benefit from a close collaboration with pre-employment risk assessors, such as industrial-organizational psychologists, to develop structured interviews that are based on actuarially derived risk assessment content. This development should include training on how to properly conduct a structured selection interview that is both thorough and non-confrontational. Studies have already shown evidence in structured interviews predicting better outcomes in the selection process than unstructured (see Cortina, Goldstein, Payne, Davison, & Gilliland, 2000; McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D., 1994). The most effective use of the polygraph test might be to assess the veracity of the answers to certain interview questions, which have actuarial relevance to risk prediction, rather than as an independent means of prediction.

Polygraph testing should not be used as an un-informed "fishing expedition." There

is little value in casting a vast net that captures information irrelevant to hiring decisions or future job performance. Nor is there incremental validity to be gained from pursuing topics that overtax the examinee's ability to properly recall, or in presenting questions that are so ambiguous that they cannot be reasonably evaluated. To maximize polygraph's effectiveness in candidate selection and to allow the highest polygraph accuracy, relevant test questions should be few in number and based on a scientific process. Such questions should be restricted to past behaviors that are empirically related to hiring decisions, and behaviors that are known performance predictors. Questions should describe behaviors or categories of behaviors that are clearly defined and do not involve attitudes, motivations, or inclinations. Questions should address time periods that can be reasonably remembered by candidates, and, time periods that are germane to hiring decisions. Significant polygraph responses to unclear or unrealistic questions may screen out potentially qualified candidates.

One very popular variant of polygraph police screening, the Law Enforcement Pre-Employment Test (Research Division Staff, 2006), uses questions such as "Are you now intentionally withholding any information about your involvement with illegal drugs?" On the surface this may seem like an ideal question but a closer look can point to potential problems. This question requires examinees to engage in the cognitive processing of their entire life's exposure and experience surrounding the concept of illegal drugs. The phrase "any information about your involvement" is very broad. Moreover, unless carefully defined the term illegal drugs could be ambiguous (e.g. Does it include a single case of taking someone else's prescribed pain medication for a migraine and/or does it include underage use of cigarettes?) The expectation for truthful persons, that every detail regarding an aspect of one's behavioral history has been disclosed, may not be realistic. Because it may not be realistic to know absolutely everything about someone's behavioral history, this type of questioning may not provide realistic and meaningfully interpretable results when an examinee admits to some historic pattern of recreational drug use.

Field examiners sometimes attempt to test the limits of an examinee's admissions (e.g., did they smoke marijuana 10 times or 11 times; 12 or 22 times), when it might be more informative to investigate the examinee's complete denial of certain behavioral activities within some recent time periods as defined by agency hiring policies and police selection research. For agencies that feel compelled to enact a rule-based decision boundary regarding behaviors such as illegal drug use, we provide several suggestions. We recommend an approach designed to elicit information to assist adjudicators in assessing and differentiating the general pattern of use (i.e., experimental, recreational, habitual, medicinal, dependent, or addicted). We suggest considering a policy that emphasizes the length of time of abstinence with regard for agency policies, as it may provide information that is of more interpretive value, compared with attempts to quantify the exact volume of use.

Topical areas regarding actuarial risk indicators, or agency hiring policies, should be explored during a structured or semi-structured pretest interview. Both structured and semi-structured interviews would employ defined content relevant to police selection, training and both methods would include open-ended topical screening questions, strategic challenges to denial, and carefully designed topical probing questions. The major difference between structured and semi-structured interview is the degree of scripting and latitude in the exact presentation of each interview question. Interview and test questions should always have focus and clarity. Relevant questions that lack clarity and operational definition should not be used. Operational definitions provide a descriptive answer to the question "what does it look like when someone does that?" The use of clearly formulated operational definition for behavioral concerns can help to reduce confusion and ambiguity and reduce erroneous and inconclusive results. Every behavior that represents a potential polygraph interview question or test question should be supported by an operational definition that provides the examiner, examinee, and risk adjudicator with a common behavioral lexicon. It is not within the realm of actuarial science or risk prediction to predict future behavior based on intention.

From an efficiency perspective, polygraph testing on topics for which information may reliably be obtained by other means (i.e., credit history, criminal records checks, etc.) should be considered redundant and wasteful of resources. Finally, it goes without saying that the topic areas must not violate law, such as the American with Disabilities Act. All of these factors should be discussed among members of the screening and assessment team (i.e., hiring officials, background investigators, industrial psychologists, polygraph examiners, attorneys representing the hiring agency) so that there can be reasonable expectations regarding the role that each can play in choosing the best candidates.

Empirically Supported Topics for Screening Polygraph Examinations

Someone who uses racial or ethnic slurs and/or engages in acts of physical and/or domestic violence as an adult seems to lack an ability to control their temper. Such actions display a lack of tolerance, which has been listed as a component of an "ideal selection battery" (Aamodt, 2004). Applicants lacking tolerance might be a considerable liability risk for any law enforcement organization as officers are often placed in situations which require them to exercise verbal and physical restraint. When it becomes necessary to use physical force, officers are trained to use only that amount of force reasonably necessary to accomplish a legitimate law enforcement objective and they should be able to deescalate their level of force once the need to use other force has passed. It would seem probable that, as adults, people prone to engage in physical acts of violence, and those intolerant of others different from themselves, may be less capable of deescalating the use of force when the circumstances dictate. Polygraph questions surrounding patterns of adult physical violence or negative ethnic or racial comments directed to others may help give evaluators better insight into veracity of answers in this area on background questionnaires.

Criminal behavior is a *prima facie* target for police screening. Applicants who have engaged in a pattern of criminal activities have shown a disregard for societal norms and appropriate social behavior. It would be

important from a polygraph question construction standpoint to have an operational definition of what constitutes target criminal behavior. An example might be selecting crimes listed in the Cormier-Lang system (Quincy et. al., 1998) for quantifying criminal history. Data suggest that applicants with a pattern of disciplinary problems or legal problems (e.g., fired from work, arrests, traffic violations, school discipline problems, etc.) should be viewed in a highly cautious manner as these were correlated with increased discipline problems and poorer performance (Aamodt, 2004). Empirical data linking such behavior to performance or risk is limited. Data collection is very difficult in this matter because applicants who disclose such transgressions are normally not hired, and therefore there is no opportunity to track their work performance and compare it with those who do not admit to past criminal behavior.

Involvement with organized crime activities represents a possible concern for some law-enforcement agencies, especially those dealing with security near portals and borders. It would be an especially desirable form of insurance for persons involved in drug trafficking or transport to have insider information or assurance that certain vehicles or activities will be overlooked in routine, random, or situational investigations. Any applicant with a history of income from or association with organized criminal activities may be deemed an inappropriate potential risk for some agencies.

Within the context of criminal behavior, adult sexual contact with minors may be of particular interest to hiring officials. Sex between adults and minors is considered especially heinous, and incompatible with law enforcement responsibilities and community trust. The scope of polygraph inquiry that may provide the best return to evaluators would be the examinee's history of sexual contact with underage persons after becoming an adult. However, other criminal sexual behaviors can be explored during the pretest interview. Caution is warranted in this area, especially when discussing romantic relationships during the applicant's young adult years. Although the age of consent in many states is 15, it is not uncommon for people to worry that their behavior may fall under the construct of this target area. Polygraph

questions can be modified to deal with these types of circumstances.

Though drug use inquiries are common in law enforcement selection, there has been little research conducted on their predictive value (Aamodt, 2004). The Equal Employment Opportunity Commission (EEOC), which enforces the Americans with Disabilities Act (ADA), allows such inquiries within certain limits. Casual and current illegal drug uses are not covered under the ADA. Earlier, we discussed the idea that there would seem to be face validity around this line of inquiry. Testing recent drug use and limiting time frames to reasonable periods is consistent with the requirements of the ADA, may add incremental validity to judgments about an applicant's pattern of illegal drug use, and would seem to increase the ability of an examinee to unambiguously answer a test question. Also, such questioning is likely to produce information that may provide better insight into present and future behavior when compared with exhaustive attempts to test the limits of historic recreational drug use as a juvenile or young adult.

One final mention of concern are questions surrounding errors and/or omissions in application forms. We should consider the wisdom and hazards of very broad questions which are intended to test the veracity of the often comprehensive forms and booklets that applicants are required to complete. One obvious complication is the known tendency for people to under-report on self-report inventories. We suggest the polygraph is a less than ideal way to verify the information provided in those booklets. One broad sweeping question attempting to verify or refute all of the information provided by an applicant would surely seem to test the limits of both the subject's memory and the capability of the polygraph test. The obvious exception is those circumstances in which inaccurate or incompleteness of information on the required disclosure forms itself represents an enforceable behavioral transgression.

Recommendations

Based on the published evidence and best practices derived from field experience,

we offer the following nine recommendations for polygraph screening of police candidates:

1. Polygraph test results should not be the sole basis for the selection or rejection of a police candidate without other information.
2. Polygraph test information and results must be kept confidential within the screening process, and not used for any purpose other than to assist in candidate selection except as provided by law. While not specifically addressed elsewhere, the reasoning for our claim seems self-explanatory. We consider the personal information divulged by applicants to be similar to that which would presumably be shared with a member of clergy, a spouse or an attorney. Absent a legal obligation or waiver to report this information, we suggest all agencies treat this with the utmost respect in regard to confidentiality.
3. Polygraph test issues should be the smallest number possible, and restricted to areas that:
 - a. Have shown empirically to correlate with officer suitability.
 - b. Are directly related to the candidate's ability to fulfill law enforcement duties.
 - c. Are restricted to overt past behaviors (not thoughts, inclinations, intentions or states of mind).
 - d. Are sufficiently recent to assure accurate recollection by the candidate.
 - e. Are sufficiently precise in definition as to avoid confusion in the mind of the candidate.
 - f. Are not adequately covered by other investigative methods already employed as part of the screening process.
4. The following polygraph test issues seem suitable for most police departments, and are either probabilistically related to poor officer conduct or relevant to other information provided by candidates:
 - a. Tolerance related issues (adult acts of physical/domestic violence, use of racial or ethnic slurs directed at others, etc.)
 - b. Criminal conduct as an adult (involvement with or income from organized crime activities may represent the primary concern for some agencies).
 - c. Illegal drug use during recent years, for agencies compelled to inquire in this area.
 - d. Formal disciplinary actions received from previous employers.
5. Examinees should be afforded the opportunity to discuss the test results with the examiner and to explain any reactions to relevant test questions.
6. Polygraph examiners should be properly trained in conducting structured and semi-structured interviews, and should remain current with changes in those fields.
7. Polygraph examiners should use polygraph protocols that include the application of the successive hurdles model to decision making. This can include decision tools other than polygraph. For example, one polygraph manufacturer, Lafayette Instrument Company (Lafayette, In.) offers a pre-employment questionnaire that can serve as an earlier phase of the successive hurdles approach. The Ryan-Personal Information Questionnaire (R-PIQ) can serve to reduce the number of polygraph examinations on the same subject by functioning to flag potential areas of concern.
8. External review of a portion of the polygraph examiners' work product each year avoids or reduces "drift" and can sustain the quality of the testing program. This is another area that was not specifically addressed earlier in the paper but bears discussion. No polygraph program should be immune from a peer review. We choose not to discuss the qualitative nor quantitative boundaries of such reviews in this paper. However, it is worth noting that there is no profession in existence in which the scope of professional behavior is beyond oversight. It goes without saying that any polygraph examiner who refuses to subject their work to peer-review may have something to hide.
9. Unless precluded by law, policy or procedure, all examinations should be recorded in their entirety. In an age in which video and audio recording technology is easily available and fully integrated into all modern field polygraph systems, there is no practical reason to forgo the advantages of a complete video

and audio recording of all polygraph examinations. It is only through complete recordings that meaningful quality assurance is possible. Frankness regarding monitoring devices helps assure the examinee the test will be conducted in a professional manner and may assist in convincing the test examinee that the examiner is being open and truthful. Brief explanation of any quality assurance program also assists in establishing a professional and trustworthy atmosphere.

Conclusions

We attempted to enlighten the reader to the potential strengths and shortcomings of applying polygraph to the public safety hiring process. Our modest attempt to provide a basic primer is hopefully seen as just that and perhaps it will encourage others to expand the dialogue around the subjects we have broached. One paper alone could not possibly cover the multitude of potentially complex issues, such as sensitivity, specificity, validity and reliability. The debate between utility and validity alone will, and should, continue. Support for utility is fueled by the popularity of polygraphy which is often based on the admissions elicited during the process. The appeal of utility however, should be tempered by concerns for validity if the test results are to be considered in the overall evaluation of an applicant.

There are a number of issues of concern surrounding the use of polygraph in the public safety employment screening process and many have not been adequately researched. Such concerns include but are certainly not limited to; sufficient training for polygraph examiners, adequate quality assurance programs, appropriate target selection for test questions, and increased precision in test result reporting, for example p-values. A paucity of empirical support directly relating to polygraph screening of multiple issues currently exists. This shortcoming alone, seems to grant enough latitude to make an oft-stated appeal for increased research in this arena and to warrant caution in interpreting test results.

We acknowledge our recommendations offer no panacea for public safety administrators. Some of the suggestions we

make could theoretically result in an increase in polygraph examiner workload. For example, the training, equipment, protocols and most importantly the quality assurance programs of local law enforcement are likely to not be equivalent to that of the federal government entities. Our recommendations include reducing the number of test questions to those that are supported from a risk-assessment consideration. We hope that by taking this approach, examination questions could be developed that shift the testing focus away from a "values-based" approach and towards one that is "empirically based." We recommend training in semi-structured or structured interviewing and feel this can be accomplished in both basic polygraph schools and at continuing educational seminars. It is likely the recommendation we make that may be perceived with reluctance by the polygraph community is the inclusion of a quality assurance program for all examiners conducting this type of testing. We realize this may place some additional burden on practicing examiners by requiring they have a portion of their work reviewed externally. We also feel we make an adequate case for the need for such a program. The recommendations for a successive-hurdles program may be seen as a policy that would tend to increase the work load of a polygraph program. We offer that such successive hurdles include options other than polygraph testing, for example the R-PIQ distributed by Lafayette Instrument Company. The recommendations we offer could tend to decrease the overall load on polygraph programs by adopting more standardized polygraph techniques and test questions and relying less on polygraph examinations and more on other means of information gathering.

While all tests tend to be regarded simplistically as pass/fail devices, it is important to remember that the overall process of law enforcement selection is a comprehensive one that involves many activities designed to elicit possible red-flags. Any candidate for whom red-flags are raised, through psychological, medical, polygraph or background investigation, should not automatically be selected out of the process. When the overall picture of an applicant becomes one in which the overall doubts outweigh the benefits, it is reasonable that the

individual is selected out of the hiring process. Polygraph testing is not intended to replace, but to add incremental validity to the police selection process. This is manifested through information gained from self-disclosure, detection of problem areas for further inquiry and deterrence of some unsuitable applicants. Since every aspect of human physiology serves multiple functions, consumers of polygraph test results should be wary of any suggestion or attempt to replace the polygraph test with a method based on a single physiological index. At present, the polygraph remains the most mature and developed form of scientific credibility or honesty testing available for use in field settings. Like all tests, the polygraph remains a useful though imperfect tool for which its ethical and competent use depends in part on its professional consumers becoming adequately informed about its capabilities, complexities and limitations.

The polygraph can be an integral part of a comprehensive hiring and review program led by psychologists, as they have the ability to integrate all parts of the assessment when making hiring recommendations. Polygraph tests can offer unique and detailed information, that when carefully integrated with other sources, can improve candidate selections. It is often difficult to find the proper weighting of the polygraph information

and results when assessing the applicant using a whole-person approach. Under-reliance on the polygraph can deny agencies critical information, particularly in the area of unreported criminal behavior. Over-reliance on the polygraph risks damaging an agency's reputation and an applicant's future. Using polygraph results as the sole basis of hiring decisions seems ill-advised. As a best practice we suggest that polygraph examiners work in collaboration with police psychologists or industrial/organizational psychologists in developing standardized structured and semi-structured interview protocols based on topical areas that are relevant to police selection. We suggest that values-based approaches be regarded with increased caution. These methods in which the interview topics and examination targets are informed by the anecdotal experience or opinions are not adequately supported by research. We further suggest the needs of law enforcement agencies and the community will be best served by emphasizing police pre-employment polygraph investigation targets that are informed by actuarial data from risk and performance prediction studies. Appropriate use of the polygraph in its decision support role can add incremental validity and contribute to overall fairness in the law enforcement personnel selection process.

References

- Aamodt, M.G. (2004). *Research in law Enforcement selection*. Boca Raton: BrownWalker Press.
- Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *Forensic Examiner*, 15(1), 6-11.
- Alder, K. (2007). *The lie detectors, the history of an American obsession*. New York: Free Press.
- American Polygraph Association (2008). American Polygraph Association Download Policies and Acts, accessed 11/20/2008 at; <http://www.polygraph.org/downloads>.
- ASTM (2002). *Standard Practices for Interpretation of Psychophysiological Detection of Deception (Polygraph) Data (E 2229-02)*. ASTM International.
- American Society of Testing and Materials International (2008). Committee E52 on Forensic Psychophysiology, <http://www.astm.org>.
- Barland, G. H. (1981). *A validation and reliability study of counterintelligence screening test*. Unpublished manuscript, Security Support Battalion, 902nd Military Intelligence Group, Fort Meade, Maryland.
- Barland, G.H. (1999). *American Polygraph Association Newsletter*, 32(3), 16-17.
- Barland, G. H., Honts, C. R., & Barger, S. D. (1989). *Studies of the accuracy of security screening polygraph examinations.*, Fort McClellan, Alabama: Department of Defense Polygraph Institute.
- Blalock, B., Nelson, R. & Handler, M., (2008). A focused polygraph technique for PCSOT and law enforcement screening programs, *The Police Polygraph Digest*, January, 2009, p. 16-31.
- Capps, M.H. & Ansley, N. (1992a). Comparison of two scoring scales. *Polygraph*, 21(1), 39-43.
- Correa, E.J., and Adams, H.E. (1981). The validity of the pre-employment polygraph examination and the effects of motivation. *Polygraph*, 10(3), 143-155.
- Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology*, 53(2), 325-351.
- Daniels, C. W.(2002), Legal aspects of polygraph admissibility in the United States, In M. Kleiner (ed.) *Handbook of Polygraph Testing* (p. 327-338), San Diego, CA, Academic Press.
- Handler, M., Nelson, R. and Blalock, B. (2008). A focused polygraph technique for PCSOT and law enforcement screening programs, *Polygraph*, 37(2), p. 100-111.
- Harris, J.C., Horner, A., & McQuarrie, D.R. (2000). *An evaluation of the criteria taught by the Department of Defense Polygraph Institute for interpreting polygraph examinations*. Johns Hopkins University, Applied Physics Laboratory. SSD-POR-POR-00-7272.
- Harwell, E. M. (2000). A comparison of 3- and 7-position scoring scales with field examinations. *Polygraph*, 29(2), 195-197.
- Honts, C. R., & Amato, S. (2007). Automation of a screening polygraph test increases accuracy. *Psychology, Crime & Law*, 13, 187-199.

- Honts, C. R. (2005). *Information gain for the psychophysiological detection of deception*. Paper presented at the American Psychology-Law Conference in La Jolla, California, USA.
- Honts, C. R., (1992). Counterintelligence scope polygraph (CSP) test found to be a poor discriminator. *Forensic Reports*, 5, 215-218.
- Honts, C. R., (1991). The emperor's new clothes: Application of polygraph tests in the American workplace. *Forensic Reports*, 4, 91-116.
- Honts, C. R. (2004). The psychophysiological detection of deception. In P. Granhag, & L. Strömwall (Eds.), *Detection of deception in forensic contexts*. London: Cambridge University Press.
- Honts, C. R., & Amato, S. L. (1999). *The automated polygraph examination: Final report*. Final report of U. S. Government Contract No. 110224-1998-MO. Applied Cognition Research Institute, Boise State University.
- Honts, C.R. & Raskin, D.C. (1988). A field study of the validity of the Directed Lie Control Question. *Journal of Police Science and Administration*, 16, 56-61.
- Horowitz, S. W., Kircher, J. C., Honts, C. R. & Raskin, D.C. (1997). The role of comparison questions in physiological detection of deception. *Psychophysiology*, 34, 108-115.
- Kircher, J. C., Woltz, D. J., Bell, B. G., & Bernhardt, P. C. (1998). *Final report: Effects of audiovisual presentations of test questions during relevant irrelevant polygraph examinations and new measures*, unpublished manuscript available from the first author.
- Kircher, J.C., Kristjansson, S.D., Gardner, M.G., & Webb, A. (2005) *Human and Computer decision-making in psychophysiological detection of deception*. Submitted to the Department of Defense Polygraph Institute, unpublished manuscript available from the first author.
- Kircher, J. C., Horowitz, S. W. & Raskin, D. C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior*, 12, 79-90.
- Kircher, J.C. and Raskin, D.C. (2002) Computer methods for the psychophysiological detection of deception, In M. Kleiner (ed.) *Handbook of Polygraph Testing* (p. 287-326), San Diego, CA, Academic Press.
- Krapohl, D. J. & Cushman, B. (2006). Comparison of evidentiary and investigative decision rules: A replication. *Polygraph*, 35(1), 55-63.
- Krapohl, D.J. (2002). The polygraph in personnel screening, In M. Kleiner (ed.) *Handbook of Polygraph Testing* (p. 217-236), San Diego, CA, Academic Press.
- Krapohl, D.J., & Stern, B.A. (2003a). Principles of multiple-issue polygraph screening: A model for applicant, post conviction offender, and counterintelligence screening. *Polygraph*, 32(4) 201-210.
- Krapohl, D.J., & Stern, B.A. (2003b). The infamous James Alphonso Frye. *Polygraph*, 32(3) 188-199.
- Krapohl, D.J., Senter, S.M., & Stern, B.A. (2005). An exploration of methods for analysis of multiple-issue relevant/irrelevant screening data. *Polygraph*, 34(1), 47-61.
- Krapohl, D.J. (2006). Validated Polygraph Techniques, *Polygraph*, 35(3), 149-155.

- Krapohl, D.J. (1998) A Comparison of 3- and 7-position Scoring Scales with Laboratory Data, *Polygraph*, 27(3), 210-218.
- Linehan, J.G. (1978). An aspect of World War II use of the polygraph. *Polygraph*, 7(3), 233-239.
- Linehan, J.G. (1990). The Oak Ridge polygraph program 1946 - 1953. *Polygraph*, 19(2), 131-138.
- Lykken, D. T. (1998). *A tremor in the blood: Uses and abuses of the lie detector*. New York: Plenum Trade.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and metaanalysis. *Journal of Applied Psychology*, 79(4), 599 – 616.
- Meehl, P.E. & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 52(3), 194-216.
- Menges, P.M. (2004). Directed lie comparison questions in polygraph examinations: history and methodology. *Polygraph*, 33(3), 131-142.
- Messig, R, & Horvath, F. (1995). A national survey of practices, policies and evaluative comments on the use of pre-employment polygraph screening in police agencies in the United States. *Polygraph*, 24(2), 57-131.
- National Research Council (2003). *The polygraph and lie detection*. Committee to Review the Scientific Evidence on the Polygraph. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Nelson, R., Handler, M., & Krapohl, D. (2008). Brute Force Comparison: A Monte Carlo Study of the Objective Scoring System version 3 (OSS-3) and Human Polygraph Scorers. *Polygraph*, 37(3), 185-215.
- Offe, H. & Offe, S. (2007). The comparison question test: Does it work and if so how? *Law and Human Behavior*, 31, 291-303.
- Podlesny, J.A., & C.M. Truslow (1993) Validity of an expanded-issue (Modified General Question) polygraph technique in a simulated distributed-crime-roles context. *Journal of Applied Psychology*, 78(5):788-797.
- Quinsey, V. L. Harris, G. T. Rice, M. E. & Cormier C. A. (1998) *Violent offenders: Appraising and managing risk*. Washington, D.C., American Psychological Association.
- Raskin, D.C. & Honts, C.R. (2002) The Comparison Question Test, In M. Kleiner (ed.) *Handbook of Polygraph Testing* (p. 1-48), San Diego, CA, Academic Press.
- Raskin, D. C., Barland, G. H., & Podlesny, J. A. (1978). *Validity and Reliability of Detection of Deception*, Contract 75-NI-99-00OI, U. S. Department of Justice, Washington, D.C.: U. S. Government Printing Office.
- Reed, S. (1994). *Briefing on the Test of Espionage and Sabotage (TES), expansion and validation study*: Retrieved on 03/13/2008 from; <http://antipolygraph.org/documents/tes-expansion-validation.shtml>
- Research Division Staff, (1997). A comparison of psychophysiological detection of deception accuracy rates obtained using the counterintelligence scope polygraph and the test for espionage and sabotage question formats. *Polygraph*, 26, 79-106.

- Research Division Staff. (1998). Psychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage. *Polygraph*, 27, 68-73.
- Research Division Staff (2006). *Federal psychophysiological detection of deception examiner handbook*. Defense Academy for Credibility Assessment (formerly the Department of Defense Polygraph Institute). Ft Jackson, SC. Retrieved 1-10-2007 from <http://www.antipolygraph.org/documents/federal-polygraph-handbook-02-10-2006.pdf>.
- Senter S.M., & Dollins. A.B. (2003). *New decision rule development: Exploration of a two-stage approach*. (DoDPI01-P-0006). Fort Jackson, SC: Department of Defense Polygraph Institute.
- Van Herk, M. (1991). Numerical evaluation: Seven point scale +/- 6 and possible alternatives; A discussion. *Polygraph*, 20(2), 70-79.
- Vrij, A. (2008). *Detecting lies and deceit: pitfalls and opportunities (2nd Ed.)*, West Sussex, England, Wiley & Sons.
- Wells, G. L., & Olson, E. A. (2002) Eyewitness identification: Information gain from incrimination and exonerating behaviors. *Journal of Experimental Psychology: Applied*, 8, 155-167.
- Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, 88, 776-784.