

Testing the Limits of Evidence Based Polygraph Practices

Raymond Nelson

Abstract

Reliance on evidence-based practices has become important in many areas of professional work, including medical, mental health and forensic practices. It is not likely that emphasis on the importance of evidence-based practices will be reduced in the future. Questions about evidence-based practices include questions about the reliability, criterion validity, and reproducibility of a polygraph test result. Polygraph methods that continue to emphasize subjectivity, arbitrary processes, procedures that cannot be tested and evaluated against an external criterion, techniques that depend heavily on un-reproducible human intuition, guesswork, and methods that cannot conform to recognizable science will ultimately come to be seen as a liability and hazard to the polygraph profession and to those communities and agencies served by the polygraph profession. This paper describes a simple analytic theory or hypothesis for the comparison question polygraph test. Discussion addresses a number of areas of field practice for which additional research is needed, including confirmatory testing, statement tests, affirmative answer tests, and testing the limits of admitted behavior.

The operational or analytic theory of the polygraph holds that differences in physiological activity are loaded onto different types of test stimuli as a function of deception or truth-telling in response to the relevant investigation target stimuli (Bell, Raskin, Honts & Kircher, 1999; Honts, 1997; Honts & Peterson, 1997; Honts & Raskin, 1988; Honts & Reavy, 2015; Kircher & Raskin, 1988; Kircher, Packard, Bell & Bernhardt, 2001; MacLaren & Krapohl, 2003; Nelson, 2014, 2015a, 2015b; Raskin & Kircher, 2014; Raskin, Kircher, Honts & Horowitz, 1988). During the examination interview an examinee will deny involvement in the behavioral issue under investigation, and is effectively denying that he or she will exhibit the kinds of differences in response that are normally expected from a guilty person who has engaged in a behavior under investigation. Differences in physiological activity can be recorded from an array of sensor technology and then combined into structural or statistical models that provide information to classify test results as deceptive or truthful based on differential salience of different types of test stimuli (Handler & Nelson, 2007; Senter, Weatherman, Krapohl & Horvath, 2010). The psychological basis for differences in physiological activity can be thought of as generally involving a combination of mental effort necessary to conceal the truth and assert a lie, emotion related to the behavioral act or the potential consequences

for the act, and conditioning to the descriptive stimulus as a result of involvement or experience in the behavioral act (Handler, Shaw & Gougler, 2010; Nelson, 2015a).

Polygraph testing is neither a deterministic (i.e., perfect and infallible) observation of deception or truth-telling nor a direct physical or linear measurement of deception or truth. Polygraph test results, like all test results, are not expected to be perfect and are fundamentally probabilistic with calculable margins of uncertainty (American Polygraph Association, 2011, Nelson & Handler, 2012, 2015; Nelson, et al., 2011), including when probabilistic results are reduced to categorical results for convenience. Probabilities associated with test results can refer to expected classification accuracy rates with groups or samples of exams, and can also refer to the statistical error or accuracy level for a single examination by comparing the examination data to empirical or mathematical reference distributions.

The National Research Council (2003) reported median accuracy as .86 using the receiver operating statistic, with in interquartile range of .81 to .91. The American Polygraph Association (2011) reported similar evidence showing the polygraph to be capable of providing accuracy that significantly exceeds chance levels. Mean accuracy for diagnostic polygraph techniques was reported as .89 with a 95%



confidence range of (.83 to .95), while accuracy of polygraph techniques used for multiple issue screening exams, for which the test questions are scored and interpreted with an assumption of independent criterion variance (i.e., it is conceivable that an examinee may have engaged in behaviors described by one or more behavioral topic are but not others), was reported as .85 with a 95% confidence interval from .77 to .93. Informed readers will be primarily interested in the more cautious lower limit of test accuracy instead of the more optimistic mean or upper limit of these confidence intervals. Comparison question polygraph accuracy rates reported in published scientific studies have been shown to be comparable to accuracy levels for scientific tests used in medicine and psychology (Crewson, 2001).

Effective use of the polygraph will depend in part on a correct understanding of a testable and falsifiable scientific hypothesis or theory that can be applied to field polygraph contexts that may include both diagnostic polygraphs used in criminal investigations, and screening polygraphs used in law enforcement employee selection, post-conviction treatment and supervision of dangerous offenders, and information/operational security. In addition to the importance of understanding the scientific basis of the polygraph, effective policies for polygraph use will also require an understanding of the limitations and potential vulnerabilities of the of the polygraph test. This report is an attempt to briefly review areas of field practice for which additional research and careful deliberation is needed to guide the formation of evidence-based field practice policies.

Discussion

Polygraph accuracy has not been established for different behavioral targets

Honts and Raskin (1988) found no differences in effect size for polygraph scores when comparing sex crime questions with other crime questions, and there is presently no basis of evidence to describe or support a hypothesis that differences in effect size are related to the topic or target issue. Although further research may produce further insights in this area, published scientific knowledge

available at the present time does not support any conclusions or assertions that criterion or classification accuracy rates may vary as a function of the behavior topic of the investigation. Neither policies nor expert opinions about test accuracy can be presently argued to be evidence-based if they attempt to make assertions about accuracy as a function of the test target issue. Instead, our present knowledge-base on polygraph accuracy is premised on research that includes a wide variety of target behaviors with a fundamental requirement that polygraph target questions that describe sexual and other crimes for which the examinee is capable of knowing the truth about his or her past behavior.

Memory testing

The polygraph test has not been studied or validated as a test of memory or intent. A general requirement for relevant questions is that they describe a behavioral issue for which the examinee can be reasonably expected to know the truth about his or her past behavior. Testing of memory functions is a psychometric endeavor that is beyond the scope of the polygraph test. Polygraph research has not investigated any operational hypothesis regarding the criterion validity of polygraph scores and polygraph results as an indicator of memory functions or memory for a specific behavior. Instead, research at this time has established the criterion validity of polygraph scores and polygraph results relative to an examinee's denial of involvement in a behavioral issue under investigation. Although polygraph test questions pertaining to memory, intent, motivation are generally not supported, the American Polygraph Association (2009a) Model Policy for Post-Conviction Sex Offender Testing section 7.1.2.G recommends that questions about memory may be used when the examinee has admitted the behavior issue under investigation. In other words, questions about memory may be used after an examinee first admits the behavioral act, and when issue of memory is then the remaining target of the investigation. Additional research is needed before any evidence-based practice policies can be formulated regarding the use of the polygraph test as a test of memory.

Testing fantasies



Discussions of polygraph testing of fantasies is most common in treatment and supervision programs for convicted sex offenders. Reduction of deviant sexual thoughts and fantasies is sometimes an important goal for patients in sex offender treatment programs. This is because sexual thoughts and fantasy are thought to play a role in the planning of sexual offenses and in victim selection. The degree to which sexual fantasies play a role in the diagnosis of sexual deviancy will remain a matter for clinical interpretation. Self-reporting is one method to gain information in this area, though it can be assumed that patients might engage in minimization and under-reporting. Polygraph testing of fantasies is intended to develop information for clinical interpretation.

Testing of fantasies, unless connected with physical actions, is outside the scope of published studies on polygraph criterion accuracy, and is outside the scope of the general practice recommendations of the American Polygraph Association. For example, the American Polygraph Association (2009a; 2009b) section 7.1.2.A-G of the Model Policy for Post-Conviction Sex Offender Testing includes recommendations that relevant questions are simple, direct, behaviorally descriptive, avoid unnecessary jargon, time-delimited, free of assumptions of guilt, and free of references to mental state or motivation. Behavior requirements such as reporting logs may be useful towards improving clinical insight into the quality and completeness of reported information, and might provide a behavioral anchor to assist in the formulation of polygraph questions that are within the scope of established field practices. Polygraph testing of behaviors associated with deviant fantasies may be a more realistic endeavor than testing fantasies that are not expressed behaviorally.

Third-party investigation or proxy confirmation

Hardy and Murphy (1996) have recommended the use of third-party relevant questions to test victims of sexual assault. They provided several examples of this type of question, including: "Are you lying when you say that man forced you to have sex?" "Did you lie when you said that man threatened you with a knife?" and "Did you lie when you lie when you

said that man hit you?" These questions make use of the logical and linguistic device "are you lying" which will appear to some readers to imply the capability of the polygraph to actually detect lies. Hardy and Murphy also provide more direct and behaviorally descriptive examples of these relevant questions: "Did that man force you to have sex?" "Did that man threaten you with a knife?" and "Did that man hit you?" Embedded in the use of these relevant stimuli is the notion that a victim of a criminal or sexual assault, who is not under investigation, can serve as a proxy for an alleged perpetrator. The logical syllogism herein is that the test result is to be taken as a basis of evidence to support a probabilistic and professional conclusion that the third-party suspect, who did not submit to polygraph testing, must have engaged in the alleged assault behavior if the results show that the examinee has not lied. However, at the present time there are no published studies to support conclusions about polygraph accuracy when testing the victim of a sexual or violent crime as a third-party proxy to determine the culpability of an alleged perpetrator who did not submit to polygraph testing.

There are several obvious problems with these suggested victim-proxy examinations, beginning with the fact that there is no clear scientific rational and no published scientific evidence to suggest that attaching polygraph sensors to record the physiological responses of an alleged victim will be capable of any meaningful correlation with the past behavior of a third-party individual who is not present at the polygraph examination. Differential salience and loading of physiological activity for relevant and comparison question stimuli has not been studied for victims of a violent or sexual assault, and it may be difficult to gain ethics committee approval to conduct studies involving these persons. However, Raskin (1987) described the results of an earlier study in which testing errors were strongly loaded on cases in which the examinee was the complainant in an alleged crime. A cautious and realistic solution will be to emphasize reliance on other investigation methods to investigate alleged perpetrators who do not submit to polygraph testing, and to use the polygraph to investigate the potential that a crime has been falsely reported when there is some basis for this concern.



Statement tests

Traditional practice is to formulate relevant question stimuli in a manner that describes the examinee's involvement in the behavioral issue under investigation (American Polygraph Association, 2009a; 2009b). Although no published studies could be located, field examiners sometimes discuss the use of a written statement exam in situations wherein the case facts do not readily conform to the standard procedure involving behaviorally descriptive relevant questions. These complications may involve the nature of the investigation target issue (i.e., sexual assault allegations not involving physical force) or an examinee who is thought to be at risk for traumatic abreaction if required to discuss the issue under investigation. The rationale for statement test is that formulation of polygraph questions regarding a written statement can somehow rectify problems associated vague or ambiguous details regarding a behavioral allegation.

Another rationale for statement exams is that they are an easier or softer way to conduct the polygraph test. This suggestion is premised on the notion that they will be experienced differently by the examinee, compared with polygraph tests that make use of behaviorally descriptive relevant questions. If the written-statement polygraph examinations are actually different in their ability to induce psychophysiological responses, then the absence of published studies means that these examinations remain experimental. This will also mean that generalization of our present knowledge about test accuracy and effectiveness is not realistic without additional research. The American Polygraph Association (2015) requires field practitioners who use experimental techniques to comply with local laws regarding the use of experimental methods, prohibits the use of experimental methods as a sole basis for professional conclusions, and advises polygraph examiners to notify the examinee and referring professional of the use of an experimental test method. A more straightforward solution will be to gather adequate unambiguous details about the behavioral incident or allegation before commencing with polygraph testing using behaviorally descriptive relevant questions as exemplified in polygraph validity research.

If written statement polygraphs can be shown in future studies to be capable of discriminating deception and truth-telling at rates similar to traditional behaviorally descriptive testing practices though with less potential discomfort to the examinee, then it may provide interesting insight into the psychological basis for responses to polygraph stimuli. If written statement exams can be shown to be effective, then there will be no benefit to the use of behaviorally descriptive relevant questions. If this turns out to be the case, then field practitioners and polygraph program administrators may begin to consider policies that require the use of this examination method for all examinations. However, some referring agents may be unsatisfied with the added sense of uncertainty surrounding test results from polygraph examinations that do not describe the examinee's involvement in the behavioral issue under investigation. More importantly, at the present time it will be unwise to abandon polygraph practices that have been shown to be effective in favor of practices that remain unstudied.

Use of comparison questions to extract information relevant to an investigation

Lundell (2014) has recommended that probable lie comparison questions are preferred over the use of directed lie comparison questions because of they provide the examiner a mechanism with which to extract additional information from the examinee outside the scope of the relevant questions. Lundell also expressed his personal belief that every examinee who passes a polygraph with truthful scores should be questioned regarding the probable lie comparison questions with the implication that the examinee may have not passed the test and may have relevant information to disclose regarding the comparison questions. Lundell further expressed his personal opinion that valuable information would remain concealed without the use of comparison questions in this manner. Said differently, Lundell's position is that experienced polygraph examiners will be unsuccessful at obtaining useful information simply by interviewing the examinee in a professional manner according to evidence-based interviewing protocols. In asserting the superiority of the probable lie comparison question over the directed lie question, Lundell has taken a po-



sition that is inconsistent with the published scientific evidence (Honts & Reavy, 2015; Horowitz, Kircher, Honts & Raskin, 1997) as summarized by Blalock, Nelson, Handler & Shaw (2012).

In taking this position, Lundell (2014) argues that comparison question topics should be selected from topical issues for which the examinee will be aware that the referring agent will view them as relevant investigation target issues. In other words, the referring agent will be interested to know the information obtained from the comparison questions, and the examinee will believe that the referring agent will want to know the test results for these questions. Lundell also assumes that all examinees are naive to the operation and scoring of the polygraph test and the actual purpose of the comparison questions as a basis for comparison and numerical scoring of the results of the relevant questions.

While there is no question that there is potentially important value in additional information, there are also potential problems in the approach argued by Lundell (2014). First, it is likely that many examinees of average or above average intelligence will have familiarized themselves with the polygraph test and the traditional probable lie comparison question prior to submitting to an exam. Although a motivated and intelligent examinee might adopt a posture of feigned cooperation with the examiner as a wiser alternative to the potential hazards of non-cooperation, authentic manipulation of these persons will be difficult or impossible to achieve.

For groups of examinees in high-contact with each other, such as police or military cadets or offenders in post-conviction programs, it is likely that they will become familiar with any polygraph procedure that routinely challenges the credibility of every examinee after every polygraph examination regardless of whether they pass or fail the test. There is little doubt that their collective experiences will lead them to a conclusion that it may be in their best interests to wait for the written test report before revealing additional information after the completion of the in-test data acquisition phase of the polygraph. If there is sufficient external pressure in the form of potential consequences for not providing ad-

ditional information these examinees will be faced with a limited range of options, which may include accepting arbitrary consequences for making no admissions, making false admissions, or developing their skill at making safe admissions to placate or manipulate the polygraph examiner and referring agent into a sense of complacent satisfaction that they are extracting additional information by routinely questioning truthful examinees.

Savvy and superficially cooperative examinees can be expected to reveal only safe pieces of information that will both appease the inquiry of the examiner while also ensuring only minimal repercussions from the referring agent, and polygraph examiners will be tempted to find this convenient and gratifying. But it is doubtful that inauthentic cooperation from a sophisticated and knowledgeable examinee will consistently lead to information of substantial value. The practice of routinely questioning truthful examinees as if they had failed the polygraph test will be much easier with persons of low average or below average intelligence, though there may be ethical considerations associated with this. The long term result of from routinely questioning truthful examinees as if they had not passed the polygraph will be that these examinees may become effectively inoculated against their next real interrogation or investigative interview.

There are also potential hazards for the examiner that could result from the attachment of a secondary, information, objective to polygraph comparison questions including the potential for contaminating and reducing the effectiveness of the polygraph test if the examiners field practices drift towards emphasizing the importance of the secondary information objective over the primary objective of providing an objective basis of comparison to evaluate the relevant questions. Although individual confidence is often high around the ability to manage potentially conflicted objectives, many professions have recognized that even experienced professionals can have great difficulty in maintaining the importance of primary objectives over secondary objectives. Indeed, prohibitions against dual relationships exist as a testament to the troublesome potential that a primary objective could become inadvertently secondary to another interest. A more viable long term field practice strate-



gy than that argued by Lundell (2014) will be for examiners to rely on evidence-based interviewing strategies that have been shown to develop useful and valuable information with less reliance on manipulation, and to reserve the practice of additional posttest interviewing for those examinees for whom the test results indicate they have more information to reveal.

Testing the limits of admitted behavior – the problem of complete truth

When an examinee makes admissions to the examiner during a pretest interview, the examiner will work with the examinee to clarify and develop the information, and will also make accommodations within the scope of the test questions. The simplest accommodation will be to remove relevant questions from the examination when the behavioral issue under investigation has become a known issue due to an examinee's admission. However, there are times in which the goals of the examination include the objective of testing the limits of the examinee's admission in attempt to identify underreporting or minimization. This is commonly accomplished by adding additional phrases to the relevant questions. For example, and examiner might add the phrase "Other than those three times, did you..." or "Other than want you already reported, did you..."

Minimal admissions

When the examinee reports a small number of incidents he or she can be asked to describe the details of each incident, and the question can be modified using the number of reported incidents. For example, if an examinee has reported and described three incidents when he or she had engaged in behavior xyz, then the question "Did you ever engage in behavior xyz?" can be changed to "Other than those three times, did you engage in behavior xyz any other times?" Examinees who have engaged in behavior xyz only three times will be cognizant of the fact that they have truthfully reported every incident of the behavior. It is expected that recorded physiological activity for these examinees will be loaded differently in response to relevant and comparison test stimuli in the same manner as other truthful examinees. Alternatively, examinees who have engaged in behavior xyz on more than the re-

ported number of occasions will be cognizant of their own deliberate and thoughtful choice to underreport and withhold information about other incidents of the behavior. Physiological activity for these examinees is expected to be loaded for relevant and comparison questions in the same manner as other deceptive examinees. Use of the number of reported incidents in the relevant stimulus question may be an important aspect of the demarcation between deception and truth-telling and the success of attempts to test the limits of admitted behavior with reasonable accuracy.

Estimating the limits of behavioral incidents

A more complex problem exists for those examinees who admit to a pattern of behavior for which the number of incidents is sufficiently numerous that he or she cannot realistically recall and describe each incident. Skillful examiners will begin the pretest interview with the goal of requiring the examinee to describe the details of each incident of behavior xyz in some systematic way, perhaps beginning with the first incident or the last incident, and proceeding until the examinee expresses to the examiner that he or she is unable to recall the details of every incident. At this point the examinee's native admission is of more value than the number of incidents, and it is often helpful to simply quote the examinee verbatim when he or she expresses that the number of incidents is too numerous to recall the details of each incident. The examinee's quoted statement can later be interpreted clinically or administratively to help understand the meaning of the behavior pattern. The examiner may proceed to develop detailed information that the examinee does recall, until the examinee expresses that he or she cannot recall any further incidents or details.

After the details of the examinee's memory are exhausted and after obtaining the examinee's statement that the number of incidents is too numerous to recall, the examiner will then assist the examinee to estimate the number of reported incidents. Skillful examiners will pursue this objective in a manner that is careful to avoid exaggeration and over-reporting. To accomplish this the examiner may simply ask the examinee to estimate the number of times he or she had engaged in behav-



ior xyz. If the examinee is hesitant to make an estimate, then the examiner may assist the examinee by providing a structured range of options for the examinee to choose from. For example, and examiner may ask: "Is it closer to five, 10, 15 or 20 times?" The goal of this interview is to engage the examinee in a realistic discussion in which the answer is provided by the examinee, thereby avoiding the potential hazard of either giving the examinee a number or range that he or she later claim was the forceful suggestion of the examiner, and also avoiding the temptation of giving the examinee a convenient range for passive and convenient vague admission or acquiescence. In this way the estimation is initiated by the examinee and not the examiner.

Interviewing the limits of admitted behavior

Regardless of the initial number of estimated incidents, the examiner can proceed to interview the limits of the examinee's estimate by using a challenging question such as this: "Is there any possibility it was more times than that number?" If the examinee provides an assertive negative answer, then the examiner has accomplished the task. If the examinee is hesitant or unassertive, then the examiner can politely but insistently question the limits of the number of reported incidents with a probing question in this manner: "What is the most possible number of times you have done that?" When the examinee responds the examiner can once again politely and professionally challenge the limits of the admission with the challenging question: "Is there any possibility it was more times than that number?" The examinee will again respond either assertively, at which point the examinee has expressed the limits of the admission, or in a hesitant manner for which the examiner can question the limits again with the same probing question: "What is the most possible number of times you have done that?" In this way the examiner can interview and question the limits of admitted behavior without leading the examinee's answer, and without encouraging exaggeration or over-reporting.

When the examinee has reported that the number of incidents is too numerous to remember the details of every incident it should be assumed that the number of reported incidents remains an imprecise estimate. Elicitation

techniques have been shown to increase information disclosure without the increase in false-confessions that results from interviewing techniques that make use of psychological manipulations (Meissner, et al., 2014). If the interview is conducted using elicitation techniques that do not depend on psychological manipulation, then the reported information is most likely not an exaggeration. Instead it can be assumed that the number of incidents has remained underreported by some unknown factor. Again, the number of reported incidents is of less practical value than the examinee's admission that the number of incidents is too numerous to remember each incident.

Formulation of relevant target questions for numerous admitted incidents

There is no expectation for precision in polygraph test results when an examinee has reported the number of incidents as too numerous to recall each incident, and there is no value in pretending precision where precision is not realistic. For this reason, inclusion of a precise numerical limit in the relevant stimulus questions should not include the estimated number of incidents after the examinee has admitted that the number of incidents has been too numerous to remember the details of each incident. Instead, the relevant stimulus question can be modified with a more general phrase, such as "Other than what you reported, did you engage in behavior xyz any other times?"

Some obvious problems exist when attempting to test the limits of patterned or repetitive behavior. Firstly, testing the limits of admitted pattern behavior will seem to depend heavily on the false hypothesis that the polygraph test can measure or detect lies *per se*. Secondly, although elicitation techniques are intended to develop information and reduce uncertainty when it is no longer possible for the examinee to know or report the correct number of incidents, studies have not fully investigated the capability of the polygraph to test the limits of admitted behavior. Polygraph examiners and referring agents sometimes express a desire to "know everything" and want to know if a person is lying about the number of reported incidents, and they are sometimes willing to interpret a truthful result



as indicative of the precision of the reported number. The suggested remedy to this dilemma is to interview the limits of behavior during the pretest, and avoid using the number of incidents in the test stimuli when the number of incidents is too great to recall the details of each incident. Thirdly, attempts to resolve the uncertainty by exaggeration or overestimation will be viewed by some as a form of false-admission or false confession in attempt to placate the polygraph examiner and polygraph test. False confessions are, in reality, a form of deception, and it is doubtful that exaggerated or falsified information will ever be of any practical value. Finally, human tendency is most often to minimize, not exaggerate, admissions of problem behavior unless there is some secondary motivation such as attention or grandiosity, or external motivation such as threats, intimidation, manipulation or other form of intense social pressure to over report, exaggerate, or make false admissions. A more realistic view will be that it will probably remain humanly impossible to know everything, and that reported information obtained from ethical and professional interviewing practices might always remain an incomplete or minimization of the actual scope or extent of an admitted behavior pattern. Fortunately, interpretation of the meaning of a behavior pattern is a task, involving the clinical or professional judgement of persons other than the polygraph examiner, for which it is rarely necessary to know everything to begin to formulate interpretations of the meaning of the information.

Most importantly, testing the limits of admitted behavior is a complex endeavor that not been addressed by existing published studies on criterion accuracy. There is no clear scientific theory or rationale to suggest why differences in physiological activity for persons who have engaged in a behavior on numerous occasions should be loaded onto different types of polygraph test stimuli in the same manner as a person who has not engaged in the same behavior. The analytic theory of the polygraph – that physiological reactions are loaded differentially for different types of test stimuli as a function of deception or truth-telling regarding the investigation target – depends on cognitive, emotional and behaviorally conditioned response potentials that exist as a function of the dichotomy between deception and truth-telling when

answering questions about a behavior under investigation. In practice, the polygraph is valuable for two potential uses: 1) to improved access to rich information that may facilitate more effective actuarial or phenomenological risk assessment, risk management and treatment planning, and 2) to provide statistically significant improvements over unassisted human lie detection when making non-deterministic (i.e., imperfect and probabilistic) classifying individuals as either deceptive or truthful. Neither of these potential benefits includes the potential to locate or establish a bright-line delimiter as to when an examinee has reported everything when the examinee has admitted to a repetitive pattern of behavior.

Although the polygraph test has been shown to discriminate deception and truth-telling at rates significantly greater than chance when investigating behavioral issues for which the examinee denies involvement (National Research Council, 2003; American Polygraph Association, 2011), at the present time there are no published studies that describe the effectiveness of the polygraph to validate or verify that an examinee has reported everything (i.e., every detail or every incident) or has been completely truthful (i.e. omitted nothing) when an examinee has admitted to a behavior for which the number of incidents is too numerous to remember the details of every incident. The goal of polygraph testing is not to pretend perfect knowledge where this is not possible, but is only to use evidence-based practices to improve the probability of effective decision making regarding the credibility of the statements of the examinee. It should not surprise any realistic-minded professional to learn that there may always remain some additional unreported information. A conservative approach to testing the limits of an examinee's admissions will first debrief the details of each incident and then include the number of incidents in the relevant stimulus question. The most conservative practice will be to formulate relevant stimulus questions only around behavioral issues for which the examinee denies involvement.

Conclusion

Reliance on evidence-based practices has become important in many areas of professional work, including medical, mental



health and forensic practices. It is not likely that emphasis on the importance of evidence-based practices will be reduced in the future. Questions about evidence-based practices include questions about the reliability, criterion validity, and reproducibility of a polygraph test result. Polygraph methods that continue to emphasize subjectivity, arbitrary processes, procedures that cannot be tested and evaluated against an external criterion, techniques that depend heavily on un-reproducible human intuition, guesswork, and methods that cannot conform to recognizable science will ultimately come to be seen as a liability and hazard to the polygraph profession and to those communities and agencies served by the polygraph profession.

Polygraph field practices must be anchored in published evidence. Establishment of field practices without scientific evidence will amount to the use of experimental practices based on mere conjecture and opinion without regard for the scientific method. Practices based on conjecture are at risk for being driven by convenience, personality, and marketing objectives instead of scientific evidence. For polygraph professionals to prevail in response to continued legal and ethical challenges to the use of the polygraph, it will be important for polygraph professionals to remain steadfastly within the established evidence-based practices that are supported by published scientific studies. In the same way that valid principles and practices of polygraph testing in general are synonymous with the principles of science, testing and decision theory, the principles and procedures of scientific polygraph testing will be similar across different polygraph practice domains. There is no known reason why the basic mechanisms of psychology, physiology and scientific testing would differ simply because of the testing context. Said differently, there is no reason that the principles of scientific polygraph will be different in the criminal investigation context, post-conviction context, civil commitment context, pre-employment context, information security, or operational security context.

The responsibility will fall first on polygraph field examiners to correctly inform and educate other professionals about the capabilities of the polygraph test. In doing this it will be important to remain within the boundaries

of scientific validity and what is known about the polygraph test, suitability for testing, applicability of testing procedures and normative reference data to population of examinees, and the theoretical and decision theoretic foundations of the polygraph test. It will also be important for polygraph professionals and others to become familiar with discussions of probabilistic error rates and the interaction of statistical error rates with prior base rates or incidence rates. In particular, it will be helpful for field polygraph practitioners and other professionals to become educated in contemporary polygraph science and evidence based practices that emphasize the differential salience of responses to different types of test stimuli and forgo continued reliance on fear-based hypotheses that have been described as false and inadequate (Handler & Nelson, 2007; Handler, Shaw & Gougler, 2010; Honts, 1997; Kahn, Nelson & Handler, 2009; National Research Council, 2003; Nelson, 2015a; Senter, Weatherman, Krapohl & Horvath, 2010). Neglecting to correctly educate others about the capabilities and limitations of the polygraph test will lead to wishful thinking that may be inconsistent with the capabilities of the polygraph or any other scientific test. This will lead to eventual frustration and aggravation when observed reality cannot be reconciled with unrealistic expectations.

Discussions about evidence-based practices and polygraph testing have already begun to include questions about whether use of the polygraph test can be shown through evidence to contribute to improved effect sizes for desired outcomes (Rosky, 2013). Outcome measures will be practical matters such as the proportion of criminal investigations that can be cleared with the polygraph compared to the use of other investigation methods, and the proportion of training failures and problem employees in law enforcement agencies compared to outcomes using other screening and selection methods, or the proportion of offenders who successfully complete treatment and supervision programs and those offenders who later commit new crimes compared to treatment, supervision and recidivism outcomes without the use of the polygraph.

The polygraph profession as a whole will need to continue to emphasize the objectivity, scientific methods, reproducibility of



analytic conclusions, and concordance with other areas of science if the profession desires to remain relevant and useful in a future that will undoubtedly include other scientific forms of lie detection and credibility assessment. Anchoring the contemporary polygraph test to polygraph practices and solutions that were sufficiently innovative to address problems and challenges 30, 40 or 50 years ago will serve only to antique the polygraph technique and will lead eventually to the polygraph profession becoming relegated to the status of a technological anachronism such as a slide rule or automobile carburetor. Reliance on false hypotheses will place the polygraph solidly in the realm of pseudoscience and may

place agencies and communities at risk for dangerous surprises that could result from decisions based on wishful thinking instead of scientific test data. Rigorous commitment to evidence-based practices and contemporary methodology (e.g., statistical algorithms, computational machines, and automated processes) will ensure that the potential contribution of the polygraph test – including both the test results and the development of rich human information – will remain a vital and useful component of the growing arsenal of scientific lie detection and credibility assessment tools available to assist investigation and risk management professionals.



References

- American Polygraph Association (2009a). Model Policy for Post-Conviction Sex Offender Testing. [Electronic version]. Retrieved 12-12-2009, from <http://www.polygraph.org>.
- American Polygraph Association (2009b). Model Policy for Law Enforcement Public Services Pre-Employment Screening Examinations. [Electronic version]. Retrieved 1/25/2012, from <http://www.polygraph.org>.
- American Polygraph Association (2011). Meta-analytic survey of validated polygraph techniques. *Polygraph*, 40(40), 203-305. [Electronic version] Retrieved August 20, 2012, from <http://www.polygraph.org/section/research-standards-apa-publications>. Polygraph Association (2015). APA Standards of Practice. [Electronic version]. Retrieved 1/5/2016 from <http://www.polygraph.org/apa-bylaws-and-standards>.
- Bell, B. G., Raskin, D. C., Honts, C. R. & Kircher, J.C. (1999). The Utah numerical scoring system. *Polygraph*, 28(1), 1-9.
- Blalock, B., Nelson, R., Handler, M. & Shaw, P. (2012). The Empirical Basis for the Use of Directed Lie Comparison Questions in Diagnostic and Screening Polygraphs. *APA Magazine*, 45-1, 36-39.
- Crewson, P. E. (2001). A comparative analysis of polygraph with other screening and diagnostic tools. Research Support Service. Report No. DoDPI01-R-0003. Reprinted in *Polygraph* 32 (57-85).
- Handler, M. & Nelson, R. (2007). Polygraph terms for the 21st Century. *Polygraph*, 36, 157-164.
- Handler, M., Shaw, P. & Gougler, M. (2010). Some Thoughts about Feelings: A Study of the Role of Cognition and Emotion in Polygraph Testing. *Polygraph*, 39, 139-154.
- Hardy, W. L. & Murphy, V.T. (1996). Victims and polygraph examinations. *Polygraph*, 25, 249-255.
- Honts, C. R. (1997, May). Is it time to reject the friendly polygraph examiner hypothesis (FPEH)? Paper presented at the American Psychological Society at the 9th annual meeting, Washington, D. C., May 23-26, 1997.
- Honts, C. R. & Peterson, C.F. (1997). Brief of the Committee of Concerned Social Scientists as Amicus Curiae United States v Scheffer. Available from the author.
- Honts, C. & Raskin, D. (1988). A Field Study of the Validity of the Directed Lie Control Question. *Journal of Police Science and Administration*, 16(1), 56-61.
- Honts, C. & Reavy, R. (2015). The comparison question polygraph test: A contrast of methods and scoring. *Physiology & Behavior*. 15-26. 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.
- Kahn, J., Nelson, R. & Handler, M. (2009). An Exploration of Emotion and Cognition during Polygraph Testing. *Polygraph*, 38, 184-197.
- Kircher, J. C., Packard, R. E., Bell, B. G., & Bernhardt, P. C. (2001). Effects of prior demonstrations of polygraph accuracy on outcomes of probable-lie and directed-lie polygraph tests (Grant No. DoDPI97-P-0016). Final report to the U. S. Department of Defense. Salt Lake City: University of Utah, Department of Educational Psychology.



- Kircher, J. C. & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology*, 73, 291-302.
- Lundell, R. G. (2014). Understanding the shortcomings of the directed die comparison questions in post-conviction sexual offender testing. *APA Magazine*, 47(4) 71-77.
- MacLaren, V. & Krapohl, D. (2003). Objective Assessment of Comparison Question Polygraphy. *Polygraph*, 32, 107-126.
- Meissner, C. A., Redlich, A. D., Michael, S. W., Evans, J. R., Camilletti, C. R., Bhatt, S. & Brandon, S. (2014). Accusatorial and information-gathering interrogation methods and their effects on true and false confessions: a meta-analytic review. *Journal of Experimental Criminology*, 10(4), 459-486.
- National Research Council (2003). *The Polygraph and Lie Detection*. National Academy of Sciences.
- Nelson, R. (2014). What does the polygraph measure? *APA Magazine* 47(2), 39-47.
- Nelson, R. (2015a). Scientific basis for polygraph testing. *Polygraph* 41(1), 21-61.
- Nelson, R. (2015b). One Word Essay (more or less): What Does the Polygraph Measure? *APA Magazine* 48(2), 29-31.
- Nelson, R., Handler, M. (2012). Using normative reference data with diagnostic exams and the Empirical Scoring System. *APA Magazine*, 45(3), 61-69.
- Nelson, R., Handler, M. (2015). Statistical reference distributions for comparison question polygraphs. *Polygraph* 44(1), 91-114.
- Nelson, R., Handler, M., Shaw, P., Gougler, M., Blalock, B., Russell, C., Cushman, B. & Oelrich, M. (2011). Using the Empirical Scoring System. *Polygraph*, 40, 67-78.
- Raskin, D. C. (1987). Methodological issues in estimating polygraph accuracy in field applications. *Canadian Journal of Behavioral Science*, 19(4), 387-404.
- Raskin, D. C. & Kircher, J. C. (2014). Validity of polygraph techniques and decision methods. In D. C. Raskin, C. R Honts, & J. C. Kircher (Eds.) *Credibility assessment*. Academic Press.
- Raskin, D., Kircher, J. C., Honts, C. R. & Horowitz, S.W. (1988). A study of the validity of polygraph examinations in criminal investigations. Final Report, National Institute of Justice, Grant No. 85-IJ-CX-0040.
- Rosky, J. W. (2013). The (F)utility of Post-Conviction Polygraph Testing. *Sexual abuse: a journal of research and treatment*, 25, 259-281.
- Senter, S., Weatherman, D., Krapohl, D. & Horvath, F. (2010). Psychological Set or Differential Salience: A Proposal for Reconciling Theory and Terminology in Polygraph Testing. *Polygraph*, 39(2), 109-117.

