Laboratory Study of Directed Lie Polygraphs with Spanish Speaking Examinees

Rodolfo Prado, Carlos Grajales, and Raymond Nelson

Abstract

A mock crime study on a Mexican population tested the effectiveness of an experimental single-sequence event-specific diagnostic polygraph technique with two relevant questions. The experimental protocol had unweighted accuracy of 87%, an 18% inconclusive rate, sensitivity of 80%, and specificity of 93%. Reliability via Kappa's statistic was .73. Study results suggest greater than chance accuracy that is consistent with other existing techniques. Although there was no observed advantage to the experimental format, these results add support for the effectiveness of single-sequence polygraph formats as similar to other formats, and for the effectiveness of polygraph examinations conducted in Spanish as similar to those conducted in English.

Keywords: *polygraph, lie-detection, diagnostic exams, directed-lie comparison questions.*

Introduction

This project was intended to estimate the diagnostic accuracy of an experimental event-specific comparison questions test (CQT) format in a laboratory setting. CQT formats have been used in both diagnostic and screening applications. Event-specific (single issue) diagnostic testing involves question formats for which the criterion status of multiple relevant questions is interpreted with the assumption of non-independence. That is all target stimuli have a shared source of variance in that they refer to a single known or alleged incident. Screening tests, conducted in the absence of a known or alleged incident, are often formulated using polygraph techniques intended to be interpreted with the assumption of that the criterion status of multiple relevant questions will vary independently. Multiple issue exams are more complex in terms of probability theory and the attentional demands placed on the subject, and have been found to be less precise than tests for which the target questions are non-independent. Because the experimental format was intended for diagnostic exams, the experimental design called for the interpretation of multiple relevant questions within each exam with the assumption of non-independence.

One characteristic of most CQT formats is that responses to multiple

Authors note:

Funding for this study was provided privately by Rodolfo Prado and the International Polygraph Studies Center, an accredited polygraph training program for which Rodolfo Prado and Raymond Nelson are affiliated as directors and instructors. The views and opinions expressed in this report are those of the authors and not the APA, or any other entity with which the authors are associated. We would like to thank the examiners that participated in this study; Julissa Gomez Varela, Mitzy Betzabeth Torres Paredes, Ricardo Reyes Monsalvo, Juan Carlos Martinez Carvajal, Miguel Angel Suarez Espitia, Raul Dirceu Garca Delgado, Juan Carlos Ortiz Sierra, Aurora Santillan Gonzalez, Daniel Arturo Saucedo Torres & Jesus Castelan Ramirez. Correspondence can be addressed to the authors at rodolfo@poligrafia.com.mx carlos@asesoriae.mx and raymond.nelson@gmail.com.

repetitions of the sequence of test stimuli are obtained in different recorded sequences. The present study was intended to investigate the use of a single sequence format for event-specific diagnostic polygraphs. The potential advantage of the single sequence format is the reduction of uncontrolled sources of variance associated with the procedures for repetition of the question sequence, with the hope that this may lead to increased precision of test results. Single-sequence CQT formats are presently used effectively for polygraph screening tests (APA, 2011; Department of Defense, 2006; Handler, Nelson & Blalock, 2008; Research Division Staff, 1995a; 1995b). We found no published information on the effectiveness of diagnostic polygraph formats that involve multiple repetitions of test stimuli in a single recorded sequence.

Polygraph is sometimes incorrectly referred to as a "lie detector" with the misunderstanding the instrument will show unique reactions when a subject lies. Polygraph instruments do not measure lies of themselves. Instead they measure differences in physiological reaction to two different types of test question stimuli: relevant questions (RQ) and comparison questions (CQ). Reaction differences have been shown to vary as a function of deception in response to RQs that describe the subject´s possible behavioral involvement in an issue of concern (APA, 2011; Honts & Reavy, 2009; Kircher & Raskin, 1988; Kircher, Kristjansson, Gardner & Webb, 2005; National Research Council, 2003). RQs generally avoid issues related to memory, intent, and motivation (APA, 2009a).

Comparison questions can be presented as a Probable Lie Comparisons (PLC) in which the subject is maneuvered into a verbal response for which is assumed the subject is probably lying, and may also be presented as Directed Lie Comparisons (DLC) question in that the subject is procedurally instructed by the examiner to answer the DLC question untruthfully (Raskin & Honts, 2002). PLCs and DLCs have been described as similarly effective, though DLCs have been suggested as less ethically questionable since their use does not involve the psychological manipulation of the subject. (Honts & Reavy, 2009; Kircher, Packard, Bell, & Bernhardt, 2001). This study involved only the use of DLC questions with the experimental test format.

The design of the experimental technique was an event-specific single-issue comparison question format consisting of two RQs, both addressing the same issue. Even when the questions are worded with small differences, the meaning of both questions was the same, and both questions described the same behavior. Included in the question sequence were two DLC questions, one neutral question and one Sacrifice Relevant Question. All questions were presented to the subjects in a single sequence, with four repetitions of each RQ. This is in contrast to most polygraph techniques, for which the examiner collects test data during several repetitions of the sequence test stimuli while stopping physiological recording and deflating the cardio cuff in between each sequence.

Possible advantages of a single-sequence diagnostic polygraph format involve the potential for increased test accuracy as a result of the reduction of sources of uncontrolled variance in the recorded data that may be introduced as a result of the traditional practices involving separate recorded sequences for each repetition of the test stimuli. Categorical test results, often reported using the terms "Deception Indicated" or "No Deception Indicated," are based on whether there is a statistically significant response by comparing a test score to cutoff scores which were previously established using tables calculated for the statistical reference distributions of guilty and innocent subjects in the normative sampling data.

Methods

Study participants were Mexican volunteers who agreed to participate after answering an advertisement placed in a well-known Mexican newspaper. Volunteer would be rewarded $200.00 Mexican pesos (MXN) for their participation. This was equivalent to 3 days of minimum-wage salary, for taking part in a study that would last from 3 to 4

hours. Interested participants called a phone number, and then received an appointment for initial screening with an examinee monitor whose role was to facilitate the participants during the study. The inclusion criteria for participation required four minimal conditions. Participants were required to:

1. be of legal age (18).
2. not be under the influence of alcohol or drugs.
3. not to be excessively tired at the time of the test and
4. not be suffering from hunger at the time of the test.

All participants received a consent sheet, informing them of the experimental requirements and details of the study. Participants were informed of their right to terminate their participation in the study at any time without penalty, and that they would be paid the advertised amount regardless of withdrawal from the study. Participants were also informed that no harm would come to the as a result of their participation. A total number of 114 volunteers participated in the study.

The participants were randomly assigned to a guilty or innocent status. Of the 114 original participants, 57 were assigned as "innocent" and 57 as "guilty." Guilty participants were asked to participate in an activity that would program their guilty status for the study. Innocent participants were instructed to engage in an activity that would ensure their innocent status regarding the RQs.

The "guilty" status letter gave the instructions that are summarized, in English (original was in Spanish), as follows:

> *Go to the dining room. In the* coffee *table you will find a red backpack, just under the table. Pour yourself a cup of coffee and "steal" $100 (Mexican pesos / MXN) that you will find inside of the backpack. After you steal the money, you will have to pass a polygraph (lie detection) test, and the issue under investigation will be the stealing of that money.*

Participants in this group were told that during the test they were to deny stealing that money, and if their involvement was not discovered during the test they would be rewarded with a bonus of $50 pesos in addition to the original $200 pesos and the opportunity to keep the $100 pesos that was taken from the backpack. They were also told that if their involvement was discovered, they would only receive the original payment for being volunteers.

In order to confirm that the instructions were understood, the monitor asked the volunteers to review the instructions. All subjects were able to adequately describe the task to the monitor who had no contact with the examiners during the project.

The "innocent" status letter gave the following instructions:

> *Go to the dining room. Pour yourself a cup of coffee and wait until someone calls you. You will then have to pass a polygraph (lie detection) test. The issue under investigation is be the stealing of certain amount of money. During that test you have only to honestly deny your involvement in the crime because you did not take any money.*

The "innocent" status subjects were also told that if they pass the test and demonstrate their innocence, they would be rewarded with a bonus of an additional $50 pesos. They also were told that, in case that the test result somehow showed them to be guilty they would receive only the original $200 pesos reward for agreeing to participate in the project.

Data were collected between December 4th and December 14, 2012 using the described convenience sample of participants from the community. Lafayette

model LX4000 polygraph instruments were used to record electrodermal activity (EDA), breathing movement, cardiovascular activity and voluntary activity. EDA was measured using skin resistance measured by standard Lafayette metal electrodes attached to the medial phalanges of the first and second fingers. Thoracic and abdominal breathing movement was measured by a standard Lafayette pneumatic tube assembly. Cardiovascular responses were recorded through the use of a Lafayette blood pressure cuff set at a pressure of 80 to 90 mmHg and placed on the participant´s calf. An activity sensor pad was placed on the subject's seat.

The study was conducted at two separate facilities: The first location – Facility A - was referred to as "House No. 3" when speaking with the participants, and consisted of an office used by the monitor to receive the volunteer participants and to assign and program their guilty or innocent states. Only one participant was scheduled at a time, and participants did not interact with each other. After programming, the study coordinator (first author) arrived to escort each participant to a second location, referred to as "Facility B," where the examiner would conduct the test. The coordinator remained blind to the guilty or innocent status of the participants until after all data were collected and analyzed.

There were 15 different examiners working in five polygraph evaluation rooms. The coordinator assigned each participant to an examiner and examination room in order of appearance. At the end of the test the coordinator accompanied the subject to the waiting room in Facility A. Each exam was evaluated by the examiner and then by a quality control reviewer. Some examinations were determined to be unusable when the quality control reviewer noted any of the following protocol violations:

- Physical illness or affliction in the subject.
- Guilty subjects not appropriately denying the robbery, i.e. confessing the crime.
- Examiners did not use the experimental format correctly.
- Interrupted or incomplete tests.
- Non-interpretable test data.

After the test and the quality control review were completed, the test results were provided to the monitor who remained in possession of information regarding the actual criterion state of each participant. Upon comparing the test result and criterion state, the monitor rewarded each participants, who then left the facility. Because the participants were dismissed after completion of the exams, none of the participants were subject to any retesting when the examination resulted in a study protocol violation.

Experimental test format

The experimental polygraph format consisted of two RQs repeated four times in a single sequence. Data would be scored using the ESS, using only the first three usable presentations of each stimulus question. The fourth presentation would be used only in the event that one of the previous stimulus presentations is unusable due to an artifact or untimely reaction, and when the results of the first three scored presentations was inconclusive.

Information regarding the questions, order and its type is summarized, in English (original in Spanish) in Table 1. The test format included: 2 neutral questions in positions 2 and 8, and repeated at positions 13 and 18; 1 sacrifice relevant question in position 3, 2 relevant questions in positions 5 and 7 (first presentation), repeated at positions 10 and 12 (second presentation), 15 and 17 (third presentation); and finally in positions 20 and 22 (fourth presentation); 3 directed lie comparison questions in positions 4, 6 and 9, repeated at 11, 14 and 16 (second presentation), and repeated again at positions 19, 21 and 23 (third presentation).

Table 1. Questions presented during the Directed Lie Diagnostic Test.

#	ID	Type	Text	Answer
1	X		The test is about to begin, please do not move and answer with yes or no to each question.	
2	1N1	N	Are we in the year 2012?	Yes
3	SR	SR	Regarding the stealing of the money, do you intend to answer truthfully each question about that?	Yes
4	1C1	C	Have you ever lied to someone who trusted you?	No
5	1R1	R	Today, did you take any amount of money from a backpack in House No. 3?	No
6	1C2	C	Have you ever stolen from someone who trusted you?	No
7	1R2	R	Today, did you take any amount of money reported stolen from a red backpack?	No
8	1N2	N	Are we in Mexico City?	Yes
9	1C3	C	Have you ever done something that would make you feel ashamed in front of your family?	No
10	2R1	R	Today, did you take any amount of money from a backpack in House No. 3?	No
11	2C1	C	Have you ever lied to someone who trusted you?	No
12	2R2	R	Today, did you take any amount of money reported stolen from a red backpack?	No
13	2N1	N	Are we in the year 2012?	Yes
14	2C2	C	Have you ever stolen from someone who trusted you?	No
15	3R1	R	Today, did you take any amount of money from a backpack in House No. 3?	No
16	2C3	C	Have you ever done something that would make you feel ashamed in front of your family?	No
17	3R2	R	Today, did you take any amount of money reported stolen from a red backpack?	No
18	2N2	N	Are we in Mexico City?	Yes
19	3C1	C	Have you ever lied to someone who trusted you?	No
20	4R1	R	Today, did you take any amount of money from a backpack in House No. 3?	No
21	3C2	C	Have you ever stolen from someone who trusted you?	No
22	4R2	R	Today, did you take any amount of money reported stolen from a red backpack?	No
23	3C3	C	Have you ever done something that would make you feel ashamed in front of your family?	No
24	XX		The test is about to end, please don´t move until I release the air in the cuff.	

Test data analysis

Because the experimental question format was intended for event-specific examinations, categorical decisions would be made at the level of the test as a whole. All examinations were scored using the Empirical Scoring System (ESS, Nelson, et al., 2011). Detailed information on the ESS, statistical reference distributions, and decision rules can be found in Nelson et al. RQs were compared to the preceding CQ. If the preceding CQ was distorted by an artifact, the RQ was compared to the closest artifact-free CQ.

The subtotal score of each relevant question *for the first three presentations* was then determined, and subtotal scores were summed for a grand total score. Initial classifications were made using the Grand Total Rule, for which the grand total cutting score was +/-4. This corresponded to probability cutting scores of .05 for both

truthful and deceptive classifications. In other words, a grand total of +4 or greater, resulted in a truthful classification, while a grand total of -4 or less resulted in a deceptive classification. Scores between -3 and +3 resulted in an inconclusive classification. For inconclusive result, the examiner would score the fourth presentation and sum the subtotal and grand total scores once again. The cutting scores of +/-4 remained the same when all four stimulus presentations were scored.

If the result remained inconclusive, or if the difference in between 2 of the subtotals is greater than 7 points, the examiner then use a Subtotal-Score-Rule (SSR). The SSR required that any one subtotal total of -6 or less would result in a deceptive classification. A subtotal score of -6 corresponded to a statistically corrected probability cutting score of .05 for deception.

Analysis

From the 114 subjects included in the study, 22 cases (19.3%) were not included in the statistical analysis due to some form of protocol violation. More than 80% of the original sample. The excluded protocol violations reduced the sample size to 92 cases, including 29 programmed guilty cases and 43 programmed innocent cases.

The analysis was performed using a combination of mathematical software, including:

- R: A language and environment for statistical computing (R Development Core Team, 2008).
- STATA v. 11.0 (Statacorp, 2009)

Results

Diagnostic accuracy was calculated excluding all inconclusive cases. This produced a sample size of 75 with, 40 of those programmed "Innocent" and 35 programmed "Guilty." Sensitivity and specificity rates were calculated including the inconclusive cases. The results are presented in Table 2. Test accuracy is the proportion of correctly classified individuals and the Error Rate is the opposite, the proportion of misclassified subjects. Sensitivity refers to the proportion of "Guilty" individuals correctly classified as deceptive. The Specificity refers to the opposite, the proportion of "Innocent" individuals correctly classified as Non-Deceptive. False Positives and False Negatives refer to the proportion of the individuals incorrectly classified as Deceptive or Non-Deceptive, respectively.

The likelihood ratio (LR) uses both the sensitivity and specificity of the test to provide an estimate of how much a test result will change the probability or odds of having a condition. The likelihood ratio for a positive result (LR+) tells you the factor by which the probability or odds of the condition increase when a test is positive (SR). The likelihood ratio for a negative result (LR-) tells you the factor by which the probability or odds of the condition decrease when a test is negative (NSR). The likelihood ratio of a deceptive result (LR+) is the ratio of the probability (likelihood) of a Deceptive result in a "Guilty" individual compared with the probability associated with an "Innocent" subject. In other terms, this number is described as the ratio: Sensitivity/(1- Specificity). The likelihood ratio of a negative test (LR-) works in the same way, yet this time describing the ratio of the probability (likelihood) of a Non-Deceptive result in a "Guilty" subject compared with the probability associated to an "Innocent" subject. This number is described as the ratio: (1-Sensitivity)/Specificity

Table 2. Accuracy of the experimental format.

DIAGNOSTIC ACCURACY	
Accuracy (Wilson's Confidence Interval)	86.667 % (77.2 %, 92.6 %)
Sensitivity (Wilson's Confidence Interval)	80.0 % (64.1 %, 90.0 %)
Specificity (Wilson's Confidence Interval)	92.5 % (80.2 %, 97.4 %)
Error Rate (Wilson's Confidence Interval)	13.3 % (7.4 %, 22.8 %)
False Positives (Wilson's Confidence Interval)	4 % (1.4 %, 11.1 %)
False Negatives (Wilson's Confidence Interval)	9.3 % (4.6 %, 18.0 %)
Likelihood Ratio (+)	10.67
Likelihood Ratio (-)	0.22

Inconclusive results are shown in Table 3, along the 95% confidence intervals. Confidence intervals were obtained through a computationally simple approach following Wilson (1927), using a refinement of the simple asymptotic method. Seventeen of the 92 cases produced inconclusive results. Over 18% of the sample participants had inconclusive examination results with the experimental format.

Table 3. Inconclusive results and confidence intervals

INCONCLUSIVE RATE	
Number of Inconclusive Results	17
Number of Inconclusive Results (Within “Innocent” Subjects)	9
Number of Inconclusive Results (Within “Guilty” Subjects)	8
Total Inconclusive Rate (Wilson's Confidence Interval)	18.5 % (11.9 %, 27.6 %)
Inconclusive Rate (Within “Innocent” Subjects) (Wilson's Confidence Interval)	18.4 % (9.9 %, 31.357%)
Inconclusive Rate (Within “Guilty” Subjects) (Wilson's Confidence Interval)	18.6 % (9.7 %, 32.6 %)

Reliability estimates for test scores with the experimental format are shown in Table 4 in the form of Cohen’s Kappa Statistic (Cohen, 1960). Use of this statistic requires an underlying assumption of homogenous ratings provided by the different raters. This assumption was accepted because all scorers and reviewers used a standardized procedure. This study consisted of multiple raters, including both the original examiners and quality control reviewers. These were reduced to two rater categories, examiner and reviewer, for this analysis.

Table 4. Diagnostic Reliability of the DLDT

DIAGNOSTIC RELIABILITY	
Cohen's Kappa Statistic (Analytic Method Confidence Interval)	0.730 (0.576 , 0.885)
Area Under ROC Curve (Analytic Method Confidence Interval)	0.862 (0.783 , 0.941)
Agreement	86.67%
Correlation	0.734

Table 5 shows the frequencies for correct and incorrect results with the guilty and innocent cases. Table 6 shows the means and standard deviations of scores for the innocent and guilty groups.

Table 5. Cross-tabulation of classification performance, excluding inconclusive cases.

		Test Result		
		DI	NDI	TOTAL
Assigned Status	Guilty	28	7	35
	Innocent	3	37	40
TOTAL		31	44	TOTAL= 75 CASES

Table 6. Descriptive Statistics of Calculated Scores

Descriptive Statistics of the Scores	
Arithmetic Mean of Scores (Within Innocents)	5.449
Standard Deviation of Scores (Within Innocents)	5.545
Arithmetic Mean of Scores (Within Guilty)	-3.256
Standard Deviation of Scores (Within Guilty)	5.350

Discussion

This project involved the study of decision accuracy of an experimental single-sequence diagnostic polygraph technique with a cohort of community participants who were randomly assigned to guilty and innocent states regarding a mock theft crime. Accuracy of the experimental technique was 87%. Inspection of the confidence intervals

reported herein and by APA (2011) indicates that the observed accuracy is consistent with the previously reported normal range of accuracy for diagnostic technique. While the observed 18% inconclusive rate was higher than reported for some other diagnostic polygraph formats, it was within to 20% ceiling that has been previously described by the APA. Observed decision accuracy for the experimental format did not satisfy the APA requirement of accuracy over .90 for evidentiary exams. Results of other studies of this experimental format, combined with the present results, will be needed to reach any conclusions about mean accuracy for this technique.

A number of limitations can be described with respect to this study, beginning with the use of a mock crime activity, and also the fact that some of the examiners had only very recently completed their academic polygraph training and had virtually no actual field experience. Ecological validity – whether the testing conditions closely approximate actual field testing conditions - of laboratory research and mock-crime activities is an important consideration, though it can sometimes easily misunderstood or mistaken for external validity. External validity is a more important concern because it refers directly to whether observed study results are likely to be observed in field practice. The practical concern is the degree to which laboratory results may overestimate effectiveness in field settings. As a matter of general understanding in most fields of scientific research, there is a clear trend that adequately conducted laboratory research has been useful and informative when attempting to understand and estimate the range of effectiveness that can be expected in field settings.

In general polygraph research in laboratory settings has resulted in lower or more conservative estimates of polygraph accuracy than research in field settings. It is possible that either examiner expertise or the motivational aspects of polygraph subjects during actual field exams may have resulted in the previously observed increases in effectiveness of field study results over those in laboratory settings. It is also possible that research conducted in field settings has been confounded by non-random sample selection methodologies that have increased accuracy of field study results over what would be observed through random sampling. Field studies, while often lacking in complete experimental control and random selection, offer the advantage of perceived ecological validity. Laboratory research offers the advantage of greater experimental control and the potential for random assignment of guilty or innocent status. Regardless of advantages and disadvantages, differences in well-designed laboratory and field study result have not been shown to be statistically significant in the past (Anderson, Lindsay, & Bushman, 1999), and there is no obvious indication that the presently observed results are an overestimation of test accuracy.

The large number of protocol violations resulting in unusable examination data deserves discussion. Nearly 20% of the examinations conducted could not be used due to heavily artifacted data that could not be interpreted and due to protocol violations on the part of the examiners. The majority of these examinations were due to the later, and we attribute this to general inexperience on the part of many of the examiners and also to the unfamiliarity of the examiners with an experimental test protocol for which the examiners had not received previous instruction or practice until the onset of this project.

There are some hypothesized advantages to a single sequence diagnostic format, beginning with the potential for increased test effectiveness that may result from reducing a source of uncontrolled response variance when starting and stopping the recording when using traditional diagnostic CQT formats. Because the experimental format did not outperform existing polygraph diagnostic format in any way, these hypothesized advantages were not observed in these study results.

Despite these obvious limitations, observed accuracy and inconclusive rates for the two RQ single sequence diagnostic format was remarkably similar to that previously

reported by the APA (2011) for other diagnostic polygraph formats with two RQs. Observed accuracy during this study was essentially identical to other formats: no better, and also no worse. It is possible that further modification and refinement of single sequence format may lead to incremental increases in testing effectiveness over existing formats.

Another interesting and important aspect of the present study is that all participants and examiners – and the first and second authors – are native Spanish speaking persons. All of the examinations were conducted in Spanish, in Mexico City. The importance of this, though not the goal of this project, is that cross-cultural decision accuracy and inconclusive rates were observed to be consistent with previously reported research results from English language studies of other diagnostic formats with two RQs. Although this study did not address the effectiveness of PLC questions, it is noteworthy that this study adds support for the assumption that accuracy and effectiveness of CQT polygraph techniques using DLC questions can remain stable across language and cultural differences. DLC questions were also previously reported as effective in screening polygraph studies conducted in native language circumstances in Iraq (Nelson, Hander & Morgan, 2012). Coupled with other advantages of DLC questions, including the transparency, non-reliance on subject naivety, and non-manipulative administration – DLC questions may continue to play an increasingly important role in both diagnostic and screening polygraphs.

Finally, we caution against the immediate use of this experimental format in field settings. It is difficult to find any sound ethical argument for the selection of an experimental protocol that does not outperform existing evidence-based practices. Until such time that further research and experience can confirm that the experimental format will provide a level of effectiveness that equal or exceeds other established methods, practitioners should remain advised that the use of experimental procedures with members of the public can be done ethically only when the examinees are satisfactorily informed and provided the opportunity to exercise informed consent regarding whether they undergo testing with an experimental method. Continued interest in single sequence polygraph formats is recommended, though additional research and development is needed before this experimental format can be used in field settings.

References

American Polygraph Association (2009a). *Model Policy for Post-conviction Sex Offender Testing*. [Electronic version] Retrieved January 25, 2012, from http://www .polygraph.org

American Polygraph Association. (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph, 40,* 194-305.

Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychology laboratory: Truth or Triviality? *Current Directions in Psychological Science, 8*, 3-9.

Cohen, J., (1960). "A coefficient of agreement for nominal scales". Educational and Psychological Measurement 20 (1): 37–46. doi:10.1177/001316446002000104

Department of Defense (2006). *Federal Psychophysiological Detection of Deception Examiner Handbook.* Retreived from http://www.antipolygraph.org/documents/federal-polygraph-handbook-02-10-2006.pdf on 3-31-2007. Reprinted in *Polygraph, 40 (1*), 2-66.

Handler, M., Nelson, R. & Blalock, B. (2008). A focused polygraph technique for PCSOT and law enforcement screening programs. *Polygraph, 37* (2), 100-111.

Honts, C. R., & Reavy, R. (2009). *Effects of Comparison Question Type and Between Test Stimulation on the Validity of Comparison Question Test. Final Progress Report on Contract No. W911Nf-07-1-0670*, submitted to the Defense Academy of Credibility Assessment (DACA). Boise State University.

Kircher, J. C., Packard, T., Bell, B. G., & Bernhardt, P. C. (2010). Effects of prior demonstrations of polygraph accuracy on outcomes of probable-lie and directed-lie polygraph tests. *Polygraph 39*, 22–67.

Kircher, J. C., Kristjansson, S. D., Gardner, M. K. & Webb, A. (2005). *Human and computer decision-making in the psychophysiological detection of deception*. University of Utah. Final report.

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.

National Research Council (2003). *The Polygraph and Lie Detection.* National Academy of Sciences Press.

Nelson, R. Handler, M. Shaw, P., Gougler, M., Blalock, B., Russell, C., Cushman, B., & Oelrich, M. (2011). Using the Empirical Scoring System, *Polygraph*, *40* (2).

Nelson, R., Handler, M. & Morgan, C. (2012). Criterion validity of the Directed Lie Screening Test and the Empirical Scoring System with inexperienced examiners and non-naive examinees in a laboratory setting. *Polygraph,* 41, (3).

R Development Core Team, (2008). R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Raskin, D. C., & Honts, C. R. (2002). The comparison question test. In M. Kleiner (Ed.), *Handbook of polygraph testing*. London: Academic (1-49).

Research Division Staff (1995). *A comparison of psychophysiological detection of deception accuracy rates obtained using the counterintelligence scope Polygraph and the test for espionage and sabotage question formats.* Report number DoDPI94-R-0008. DTIC AD Number A319333. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in *Polygraph, 26* (2), 79-106.

Research Division Staff (1995). *Psychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage.* DoDPI94-R-0009. DTIC AD Number A330774. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in *Polygraph, 27*, (3), 171-180.

StataCorp. (2009). Stata Statistical Software: Release 11. College Station, TX: StataCorp LP

Wilson, E. B. "Probable Inference, the Law of Succession, and Statistical Inference," *Journal of the American Statistical Association, 22*, 209-212 (1927).