Polygraph

VOLUME 29 2000 NUMBER 2

Contents

Published Quarterly

P.O. Box 8037, Chattanooga, Tennessee 37414-0037

We Remember 1952

Raymond J. Weir, Jr., Walter F. Atwood, Norman Ansley

Introduction

We, the authors, were invited by the APA to record our responses to a set of questions to be posed during a videotaped interview. We have declined: we doubt that anyone would watch it. Also, over the years the technology will change so much that we believe that someday soon no one will be able to playback the recording. Instead, we offer this trilogy of personal recollections, covering what we think are important dates and events in the history of our profession.

Each of us was trained at the Keeler Polygraph Institute in 1951. By 1952 our daily testing had settled into a routine and we had the time to look at our profession. So we have taken 1952 as a point in time for our recollections, but have left the year where appropriate.

We have divided our recollections into three groups:

Applications of the Polygraph
Polygraph Training and Instrumentation
Examiners and Their Organizations

We have all contributed to each paper, but Ray Weir is the senior author of the paper on Applications, Walt Atwood on Training and Instruments, and Norm Ansley on Examiners and Their Organizations.

The authors are Life Members of the APA. Mr. Weir and Mr. Atwood are past presidents of the APA and Mr. Ansley was, for 25 years, Editor of APA publications. We regret to announce that Raymond J. Weir, Jr. passed away on April 20, 2000.

We Remember 1952 – Applications of the Polygraph

Raymond J. Weir, Jr., Walter F. Atwood, Norman Ansley

The Korean War

We three were recruited for service as polygraph examiners in the Armed Forces Security Agency (AFSA) by Major Fred Hazard. AFSA was expanding rapidly but all employees and military assignees needed Top Secret access, and such access was granted only on the basis of a national agency check and a background investigation. Hundreds of civilians had been hired but could not be put to work without a clearance, and that was taking from six to eighteen months. The concept of a national agency check and a polygraph examination for an interim clearance was proposed and accepted. Finding examiners was another matter. None were unemployed, and of those who might be available, many lacked basic training. This fact soon became evident. Some of those hired as examiners were incompetent, and not kept. Others were nearly so, and stayed only a short while. We worked seven days a week, until we had processed all those hired and awaiting clearance. Then we settled into a system of testing shortly after an employee entered on board. Later, we began to test applicants. AFSA was partly staffed with military personnel from the Army, Navy and Air Force. The chief of the polygraph branch was Major Donald F. Hermes, who was a trained examiner, and had been the business manager at Keeler Institute, and the Commanding Officer of a Reserve CIC Detachment. Second Lieutenant Lincoln M. Zonn was briefly assigned to AFSA as an examiner in July 1951. Most of the examiners were civilians, and Miss E.A. Boulanger had been told she would be the chief of polygraph when she was sent to the Keeler Institute for training in early 1951. It didn't work out. Miss Boulanger was from New England, spoke French, had a masters degree in psychology, and was probably the first woman to serve as a full-time examiner in the U.S. Government.

In 1957, though outside the scope of this paper, Miss Boulanger solved the mysterious pink ribbon evidence, a ribbon received in connection with a personnel matter from the Boston Field Office. She recognized the marking on the ribbon as Gregg shorthand and the language as French.[1]

We don't know who decided to use polygraph testing in the clearance processing at AFSA. The idea had been around since World War I when Lieutenant William W. Marston of the U.S. Army trained a group of psychologists at Camp Greenleaf, N.C. to give the tests. The Armistice put an end to the project before they went to work. Using the polygraph was considered in 1941, and briefly described by Wolfle (1941, 1946). The polygraph had also been used at the end of World War II to screen German POWs for the post-war assignment to police positions (Linehan 1978). However, an immediate precedent was an on-going program of testing employees of the Atomic Energy Commission facility at Oak Ridge, Tennessee, by a private firm owned by Russell Chatham. His company examiners also tested employees of the Diamond Ordnance Fuze Laboratory in Maryland[2] and he had two employees testing at AFSA for three months in 1951, Rex Ramsey and Arnold Cohen. Chatham described his operation at Oak Ridge on a number of instances, including forums at the University of Tennessee and New York University, both in 1952. However, the Oak Ridge operation was under attack by members of Congress and the press, and it was abolished. Perhaps the most influential program in developing the polygraph program at AFSA was the program at the Central Intelligence Agency. The CIA program's first director was Cleve Backster. In late 1952, the AFSA became the National Security Agency.

[1] It was the announcement of the engagement of Norm Ansley, then in the Boston Field Office, to Miss Nancy Pearson.

[2] The Diamond Ordnance Fuze Laboratory became the Applied Physics Laboratory of Johns Hopkins University.

Agencies Using the Polygraph in 1952

Federal

Army Counter Intelligence Corps
Army Criminal Investigation Division
Air Force Office of Special Investigations
Office of Naval Intelligence
National Security Agency
Central Intelligence Agency
Federal Bureau of Investigation (rarely)
Postal Inspectors (B & W meters)
Atomic Energy Commission
Air Force Intelligence
U.S. Secret Service
Diamond Ordnance Fuze Laboratory

State Police or State Crime Laboratory

Illinois
Michigan
Minnesota
Nebraska
New York
North Carolina
Pennsylvania
Rhode Island
Texas
West Virginia
Wisconsin
District of Columbia

Some of the Cities

Berkeley
Chicago
Evanston
Detroit
Los Angeles
San Jose/Santa Clara County
Saint Louis, MO
Seattle
Toledo
Waco
Wichita

Laws and Regulations

In 1952 there were no federal or state laws on polygraph testing. The *Frye* decision stood as a general bar to polygraph evidence in criminal trials. Although the Army had a regulation on polygraph testing, regulations were not yet common. We don't know how many examiners there were to regulate in 1952, and the International Society for Detection of Deception [ISDD] didn't say how many members they had. By 1953, the ISDD had 132 members and 10 applicants pending. Regulators had not yet paid any attention to the field, nor had the US Supreme Court ruled on polygraph results as evidence.

There were some exceptions to *Frye* in trial courts of various jurisdictions. Professor Fred E. Inbau had footnoted some of them in his 1935 article on the Wisconsin case, *State v. Loniello and Grignano*[3]. The case was significant because of the stipulation to admit the polygraph test results, before the test was given, and the judge's instruction to the jury. Despite the exceptions, the general view in 1952 was that polygraph test results were inadmissible but confessions following tests were admissible.

Applications in 1952

It appears that over half of the ISDD members in 1952 used polygraph examinations in the investigation of crime, civilian or military. Other applications included civilian testing for employment and continued employment, federal screening for access to classified information, testing of juvenile delinquents, testing for insurance fraud, paternity disputes, and testing for the defense counsel. Because of the war in Korea there was a significant increase in federal testing for classified access and in counterintelligence cases. The Communist fellow-travelers attacked this application of polygraph testing through the press and in Congress. Except for the Oak Ridge program, they failed to stop counterintelligence and security screening.

[3] Inbau, Fred E. (1935). Detection of deception technique admitted as evidence. *Journal of the American Institute of Criminal Law and Criminology*, 26 (2), 262-270.

We Remember 1952 – Polygraph Training and Instrumentation

Walter F. Atwood, Norman Ansley & Raymond J. Weir, Jr.

Instruments in 1952

The authors primarily used the Keeler model 302-C, manufactured by Associated Research, Inc. of Chicago. It was a 3-channel instrument in a large metal toolcase-like container. In addition to the 110 volt, 60-cycle current, it had a battery pack that served the electrodermal unit. It cost $1,192.50 and weighed 46 pounds. The kymograph had speeds 6-inches and 12-inches per minute, the faster necessary where there was 25-cycle current. The galvanometer worked on both conventional and self-centering circuits. The Keelers saw a lot of service, but also spent quite a bit of time on the workbench.

We had one 3-channel instrument manufactured by C.H. Stoelting Company of Chicago. The case was covered by material that looked like alligator hide. Five different batteries were required for the EDA unit. Electrodermal units were not commonly used in 1952, and 2-channel instruments like the Keeler models 304 and 305 were popular. An option on the model 305 was a spring-wound kymograph, which eliminated the need for electrical power.

Lafayette Instrument Company made polygraphs but had not yet entered the law enforcement market. Several police departments had 2-channel Berkeley psychographs manufactured in California by C.D. Lee & Sons. Not unlike today's concern about voice stress equipment, the concern in 1952 was about the use of recording and non-recording galvanometers for detecting deception. They were manufactured by B & W Associates, Fordham University, Lafayette Instrument Co., C.H. Stoelting, and Thompson Metrigraph.

The cardiograph and respiration channels of instruments in 1952 were pneumatic in operation. All pen systems had trouble with pen ink flow. There was jamming of chart paper by sprocket drives.

The current instruments require less service and produce better patterns. We know there are some dinosaurs who disagree, and continue to use instruments with mechanical tambors.

Polygraph Training in 1952

Although there were two polygraph schools in operation, we think it probable that many if not a majority of the examiners in 1952 were preceptor trained. An outline of the two courses is attached.

The Keeler Polygraph Institute offered a six-week course of full-time study and practice at its facilities in Chicago. A certificate was given only upon completion of the course and a certified report of 150 cases.

In July, 1951, the Provost Marshal General's School at Camp Gordon, Georgia began a ten-week course of instruction for Army CID agents. The course also trained examiners for the Metropolitan Police of Washington, DC and other agencies. However, Army CIC, NSA and most other federal agencies sent their agents to the Keeler Institute for training.

John E. Reid & Associates offered a six-month apprenticeship course. Many of those trained by Reid stayed on as staff examiners. While the Keeler and Army courses taught R-I and Peak of Tension, the Reid School taught Reid's comparison [control] question technique and Peak of Tension.

For details of commonly used instruments see Ansley, N. & Weir, R.J. Jr. Selected Papers on the Polygraph. Washington: Board of Polygraph Examiners, 1956, pp. 22-41.

A few examiners were trained in college courses. Dr. V.A. Leonard taught such a course at Washington State College and Dr. Douglas Kelley taught some students at the University of California at Berkeley. There was also a private school that operated briefly in Washington, DC. Called National Lie Detection, Inc. it was at 1919 K Street N.W. The firm offered basic training and polygraph services. It was staffed by Colonel Ralph Pierce, Sam Souza, Leonard Harrelson, and (fnu) Van Cleve*. Founded in May 1951, it folded in 1952. The C.H. Stoelting Company began a correspondence course in 1952 that offered basic polygraph training. The number trained, if any, is unknown.

There were two advanced training opportunities in 1952. There was the New York Conference on Criminal Interrogation and Lie Detection at Vanderbilt Hall, New York University Law Center, on November 8, 1952, featuring 14 speakers. A transcript of the proceedings was published by the Board of Polygraph Examiners and is an Appendix to Ansley & Weir, Selected Papers on the Polygraph. The International Society of Detection of Deception (ISDD) held a seminar from Thursday morning, September 11th, through Sunday evening, September 14th at the Maryland Hotel in Chicago. No transcript is available but it was a good program.

The Keeler Polygraph Institute

In 1952, the course at the Keeler Polygraph Institute was revised, and certificates were awarded only after the examiner had completed 150 cases. This requirement was not in effect when Atwood, Ansley and Weir attended in 1951. When Norm Ansley attended KPI in February and March 1951, Jack Harrison was director, polygraph topics were taught by Harrison, Cleve Backster, and Al Breitzman. Richard Inman of Associated Research taught instrumentation. Dr. S. Guten taught psychology, Dr. LeMoyne Snyder taught homicide investigation, Dr. William Wisedorf taught psychiatry, Prof. Robert A. Scott taught law, Judge Steed also taught law, staff examiners were Austin Souza, Gerry C. Forster, and Charlie Wilson. Carole Greene was the secretary. After the 1952 reorganization of KPI, Albert Breitzman became Director, and his staff included Dr. Herbert Lyle, Professor Robert Scott of Michigan State College, Duke Mattei, Professor Walters of Northwestern University and others. The 1952 course breakdown was:

Interrogation	10 hours
Chart Interpretation	18
Psychology	14
Physiology	26
Polygraph Technique	13
Technical Aspects	26
Practice	37
Misc. instruction	20
Misc. school procedures	16
Total	180

The Provost Marshal General's School

The ten-week polygraph course at the PMGS, Camp Gordon, Georgia, provided training of examiners for the Army CID, the Metropolitan Police of Washington, DC, and possibly for others. The course, which began in July 1951, was under the direction of Captain C.N. Joseph. His instructors were Captain James A. Whicanack and CWO Mervin Cumpson. Captain Joseph was replaced by Major Jack B. Richmond.

The 1952 course breakdown was:

Introduction	6 hours
Mechanical Training	32
Psychology	12
Physiology	20
Technique	60
Practice	166
Total	296

*Probably Robert E. Van Cleve.

We Remember 1952 – Examiners and Their Organizations

Norman Ansley, Raymond J. Weir, Jr., Walter F. Atwood

Organizations

There were three organizations of note in 1952: The Board of Polygraph Examiners, the International Society for Detection of Deception, and the Court of Last Resort.

Board of Polygraph Examiners

The Board of Polygraph Examiners was founded in Washington, DC in early 1952. The address was Box 7599 Benjamin Franklin Station, Washington 4, DC. Membership required full-time assignment as an examiner and formal basic training. There were thirty to forty members. Although the BPE still exists, the BPE was most active from 1952 to 1956 when most of the members joined the American Academy of Polygraph Examiners (AAPE). In 1956 Fred Inbau was President of the AAPE and John Reid, Vice President. Ray Weir and Walt Atwood who had served as presidents of the BPE, later served as officers of the AAPE. Norm Ansley, who had been secretary of the BPE also became an officer of the AAPE. The BPE and AAPE had nearly identical standards. Despite the merger, members of the BPE continue to meet occasionally, and even bestow membership on worthy candidates, a lifetime honor. During its operational phase in the early 1950s the BPE set standards for instruments, and issued a seal of approval for those meeting the standards. In 1953 the Board published a catalog, describing ten instruments that met our standards, seven which did not, and four instruments no longer in production. The BPE published Bulletins which covered techniques, ethics, legal decisions, book reviews, bibliographies, a glossary, and in 1952 the complete transcript of a conference at New York University. Many of the BPE papers, including the NYU transcript were published as a book (still available from University Microfilms International, 300 Zeeb Rd., Ann Arbor, MI 48106, No. PB2 OP69619). The BPE was not one of the organizations that merged, under guidance of Walt Atwood, to form the APA in 1966.

The International Society for the Detection of Deception

In 1952 the ISDD was the largest organization of examiners. The officers were President Herbert P. Lyle, M.D. of Cincinnati, Vice President Nathan W. Heller, attorney in Milwaukee, and Secretary-Treasurer C.B. Hanscom, Crime Lab, University of Minnesota. Chairman of the Board was Alex Gregory of Detroit, and members of the Board were Charles M. Wilson of the Wisconsin State Crime Lab, Colonel Ralph W. Pierce of Arlington, Virginia, and Freeman B. Ramer of the Pennsylvania State Police Crime Laboratory. Add some of the ISDD committee members and you had a list of many of the interesting or influential examiners of that time: Ralph G. Orcutt, Douglas Kelley, M.D., James F. Inman, Albert Breitzmann, Albert Langtry, Guido L. Mattei, F.W. Baleiko, Starke Hathaway, Ph.D., Clyde Dailey, and Leland W. Gillespie.

The ISDD held their fourth annual seminar in the fall in Chicago. We attended and it was a good seminar. In 1952 the ISDD supported the survey on the uses of the polygraph by Dean William H. Wicker of the University of Tennessee, who produced a paper on ethics, published articles on cases and techniques, noted criticism of federal screening with the polygraph, and produced an issue of the Bulletin on arson cases.

In September 1953 the ISDD reported they had 132 members and ten applications pending. Sixteen were in the military service. There were 21 in Illinois. ISDD members in Washington, DC were Cleve Backster, Lloyd Furr, Colonel Calvin Goddard, Leonard Harrelson, James K. McCarty, Ralph G. Orcutt, and Colonel Ralph Pierce.

Some members did not like the organization's name, so the ISDD became the Academy for Scientific Interrogation, and later merged with other groups to form the APA.

The Court of Last Resort

The Court of Last Resort was well known for its work in saving from execution or long prison sentences persons who had been convicted of a crime but were innocent. In a few cases the Court became involved before trial. Polygraph testing was part of the investigative process, resulting in favorable press for the profession. In 1952 Alex Gregory was the Court's examiner, having replaced Keeler who died shortly after the Court was formed. In 1952 Gregory was also Chairman of the Board of ISDD. The Court was founded and largely funded by Erle Stanley Gardner, and all members volunteered their time. Others in the Court were criminologist and attorney LeMoyne Snyder, M.D., detective Raymond Schindler, handwriting expert Clark Sellers, attorney Marshall Houts, and Harry Steeger. Houts states the Court investigated about six hundred murder cases. Among its earliest and most famous was the murder of Sir Harry Oakes. The murder on Nassau involved the Duke of Windsor, the mob, and a number of society figures including Alfred de Marigny who was found not guilty. Keeler tested him after the trial and found him not deceptive, and the test is described differently in three books on the case. The description by Marshall Houts in his *Kings X* (William Morrow & Co., 1972) is probably most accurate, as he included the irrelevant questions, and as counsel for the Court for many years, he was familiar with testing. The other books are James Leasor's *Who Killed Sir Harry Oakes?* (Houghton Mifflin 1983), and a book by the defendant Alfred de Marigny who wrote *A Conspiracy of Crowns* (Bantam Press 1990). For more information on the Court's members and cases see Erle Stanley Gardner's *The Court of Last Resort* (William Sloane 1952 and Pocket Books, 1954 revised). See also Dorothy B. Hughes, Erle Stanley Gardner, *The Case of the Real Perry Mason* (William Morrow & Co. 1978).

Influential People in 1952

Cleve Backster founded the CIA polygraph program, and was a founding member of ISDD and BPE. He developed a CQT and founded a school.

Albert Breitzman of the Evanston Police Department. Instructor and Director of KPI. Active in ISDD.

Russell Chatham had contracts to conduct security screening at the AEC plant at Oak Ridge, Tennessee and at the Diamond Ordnance Fuze Laboratory in Maryland. Funded a survey on validity and a symposium at the University of Tennessee. Member of ISDD.

Alex Gregory – Private practice in Detroit, examiner for the Court of Last Resort, and an officer of the ISDD.

C.B. Hanscom – State examiner at the University of Minnesota. Served as Secretary and other positions in the ISDD and organizations that followed the ISDD.

Fred E. Inbau – Author of a book on polygraph testing and interrogation that had gone through two editions, and would go through a third in 1953. Worked closely with John E. Reid.

Herbert P. Lyle, M.D. – President of the ISDD in 1952 and an instructor at KPI.

Clarence D. Lee – At Berkeley Police Department worked with Larson on testing in the 1920s. Retired, he manufactured the Berkeley Psychograph. Author of a book on polygraph testing.

Retired Colonel Ralph W. Pierce – On ISDD board, taught at KPI and at another school in Washington, DC in 1952.

LeMoyne Snyder, M.D. – Active in ISDD. Not an examiner but nationally known for homicide investigation. An important member of the Court of Last Resort.

The Polygraph Literature in 1952

A number of books had been published, but some were little known and difficult to obtain. Hugo Munsterberg's *On the Witness Stand* was published in 1908 and could be found in used book stores if you were lucky. The same was true of John Larson's *Lying and its Detection* (1932) and William Marston's *The Lie Detector Test* (1938). Harold

Mulbar's 1951 book *Interrogation* had two good chapters on polygraph testing. C.D. Lee's book *The Instrumental Detection of Deception* was published in 1953, but the text of the book without the pictures had been published by Lee in 1943 as *The Instruction Manual for the Berkeley Psychograph.* Christian A. Ruckmick's 1936 book on *The Psychology of Feeling and Emotion* had considerable information on detection of deception, including research. There were three editions of *Lie Detection and Criminal Interrogation* by Fred Inbau. The first edition in 1943 was strictly relevant-irrelevant and peak of tension techniques. The second edition in 1948 included an "Alternative Test Procedure – The Reid Technique." The Third Edition, published in 1953 had John E. Reid as co-author

Organizations

Two organizations published polygraph material - The International Society for the Detection of Deception (ISDD) and the Board of Polygraph Examiners (BPE). The ISDD Bulletin was irregular in production, but lengthy. It had news about the Society, clippings on the polygraph in the news, and an occasional article on a specific topic. The June 1952 issue was devoted to arson investigation. The BPE publications were also irregular, and all were on specific topics. BPE and ISDD mimeographs for reproduction.

Russell B. Chatham financed a survey and the preparation of papers in 1952, and the results were published in the *Tennessee Law Review* (February 1953) v. 22 as The Polygraphic Truth Test, A Symposium.

The Research Literature Available in 1952

We suspect the influence was generally minimal. Winter's 1936 study may have ended interest in the word-association test. The Ellson report on research at Indiana University issued in September 1952 was so disappointing to its sponsors that further work was halted. The Indiana study was largely theoretical and exploratory and did not provide practical improvements sought by the sponsors.

Publication did not mean that polygraph examiners read it, or even knew of it. Only one journal was seen by a number of examiners, and that was the *Journal of Criminology, Criminal Law, and Police Science*. As we recall, there was little interest in academic research in 1952.

Research Reports Available in 1952

Field Studies – Real Cases

Bitterman, M.E. & Marcuse, F.L. (1947). Cardiovascular responses of innocent persons to criminal interrogation. *American Journal of Psychology.* 60. 407-412.

Larson, John A. (1923). The cardio-pneumo-psychogram in deception. *Journal of Experimental Psychology.* 6 (6). 420-454.

Lyon, Verne W. (1936). Deception tests with juvenile delinquents. *Journal of Genetic Psychology.* 48 (3). 494-497.

Marston, W.M. (1921). Psychological possibilities in deception tests. *Journal of Criminal Law and Criminology.* 11 (4). 551-570. Reprinted in *Polygraph* (1985). 14 (4). 321-339.

Reid, J.E. (1945). Simulated blood pressure responses in lie detector tests and a method for their detection. *American Journal of Police Science*. 36 (1) 201-214. Reprinted in *Polygraph* (1982). 11 (1). 22-36.

Reid, J.E. (1947). A revised questioning technique in lie-detector tests. *Journal of Criminal Law and Criminology.* 37 (6). 542-547. Reprinted in *Polygraph* (1982). 11 (1). 17-21.

Summers, W.G. (1938). The electronic pathometer. Proceedings of the International Association of Chiefs of Police. Washington, DC. P. 142.

Winter, J.E. (1936). A comparison of the cardio-pneumo-psychograph and association methods in the detection of lying in cases of theft among college students. *Journal of Applied Psychology.* 20. 243-248.

Laboratory Studies-Simulated Cases

Benussi, V. (1914). Die atmung asymptome der luge (The respiratory symptoms of lying). *Archiv fur die Gestamte Psychologie*. 31. 244-273. Translated and reprinted in *Polygraph* (1975). 4 (1). 52-76.

Berrien, F.K. & Huntington, G.H. (1943). Exploratory study of pupillary responses during deception. *Journal of Experimental Psychology.* 32. 443-449.

Burtt, H.E. (1921). The inspiration/expiration ratio during truth and falsehood. *Journal of Experimental Psychology.* 4 (1). 1-21

Crane, H.W. (1915). A study in association reaction and reaction time with an attempted application of results in determining the presence of guilty knowledge. *Psychological Monographs*. 18 (4). 1-72.

Ellson, D.G. (15 September 1952). A report of research on detection of deception performed by Indiana University under contract no. N6onr-18011 with the Office of Naval Research.

Geldreich, E.W. (1941). Studies of the galvanic skin response as a deception indicator. *Transactions of the Kansas Academy of Science*. 44. 346-351.

Geldreich, E.W. (1942). Further studies in the use of the galvanic skin response as a deception indicator. *Transactions of the Kansas Academy of Science*. 45. 279-284.

Keeler, L. (1930). A method for detecting deception. *American Journal of Police Science*. 1 (1). 38-51.

MacNitt, R.D. (1942). In defense of the electrodermal response and cardiac amplitude as measures of deception. *Journal of Criminal Law, Criminology and Police Science*. 33. 266-275.

Marston, W.M. (1917). Systolic blood pressure symptoms of deception. *Journal of Experimental Psychology.* 11 (2). 117-163. Reprinted in *Polygraph* (1989). 14 (4). 289-320.

Rourke, F.L. & Kubis, J.F. (1948). Studies in detection of deception: Determination of guilt or innocence for psychogalvanic (PGR) records of delinquents and non-delinquents. *American Psychologist*. 3. 255.

A Critical Analysis of Honts' Study: The Discussion (Stimulation) of Comparison Questions

James Allan Matte

Abstract

Honts (1999) selected eleven laboratory studies that included a discussion of questions and/or the stimulation of comparison questions between the repetitions of the question list for comparison with eight laboratory studies where comparison questions were not discussed between repetitions. According to Honts' analysis of the results of those studies, the error rate was significantly reduced where questions were reviewed between repetitions, especially with guilty subjects where the error rate was reduced by 54%. Honts asserts that these results clearly support the review of questions between charts, and that the attacks against its practice by Abrams in several court cases have had a negative impact on the admissibility of polygraph examinations in United States courts of law. A critical analysis of Honts' study reveals selective scholarship and a seriously flawed research methodology, which call into question the conclusions of his study.

Key words: Comparison question, directed-lie comparison question, false negative, probable-lie comparison question, psychological set, zone comparison indication-remedy, tri-zone reaction combinations.

The Honts (1999) article regarding the inter-chart discussion of comparison questions appears to be an attempt to justify a procedure that lacks even face validity. The discussion of comparison questions alone, or their stimulation with mere inquiry about the relevant questions between charts, must have a psychological effect on the examinee whether innocent or guilty of the offense for which he or she is being polygraphed. The effect of this procedure on the psychological set of the examinee is not selective of the examinee's guilt or innocence. The results of a recently completed study (Matte & Reuss, 1999) show that the discussion of comparison questions between the charts or repetitions for the guilty examinee can have the effect of increasing the examinee's apprehension toward the directed-lie comparison questions (DLCQ), a comparison question normally used by Honts. The resultant increased apprehension of the guilty examinee to the DLCQ shifts the examinee's psychological set from the relevant questions to the DLCQs. This shift may evoke a correspondingly greater physiological arousal to the DLCQs than the relevant questions, resulting in a false negative.

This author does not believe that any competent polygraphist would subscribe to the practice of discussing the relevant (crime) questions alone between charts, or doing so with a mere inquiry about the comparison questions, for the obvious reason that it could reorient the innocent examinee's psychological set from the comparison questions to the relevant questions, inviting a false positive result. Conversely the opposite is also true (a possible exception, Combination H of Backster Rule 4, is discussed below): a false negative could result from discussing only the comparison questions between charts (Honts & Raskin 1988; Horowitz, Kircher, Honts & Raskin 1997), discussion of both the relevant and comparison questions, but placing more emphasis on the probable-lie comparison questions (PLCQs) (i.e. Honts 1999; Honts & Gordon 1999; State of Montana v. Gordon, Jefferson County Court, Case No. D.C. 97-154), can also reorient the guilty examinee's

The author is a member of APA and a regular contributor to this journal. Reprint requests should be addressed to Dr. James Matte, Matte Polygraph Services, 43 Brookside Drive, Williamsville, NY 14221-6915, or to his e-mail address, JamesAllanMatte@mattepolygraph.com.

psychological set, inviting a false negative result. But in the case of the DLCQ, the probability of a false negative would be expected to be significantly greater, because examinees better understand how the reactions to it will be used for comparison against the reactions to the relevant questions.

Interestingly, both studies by Honts & Raskin (1988) and Horowitz, Kircher, Honts & Raskin (1997) indicate that only the comparison questions (PLCQs and DLCQ) were reviewed between each chart. There is no mention in either of the aforesaid studies of any review of the relevant questions between each chart.

In his article, Honts states, "The present results could be criticized because they are from laboratory studies. However, it is very difficult to imagine that false negative outcomes are easier to produce in real case rather than in simulations. Certainly, most scientists and polygraph examiners would agree that the relevant issues of nearly all real cases are more salient than the relevant issues of laboratory simulations."

The case Honts makes that the discussion of comparison questions would have less impact on the guilty examinee in a real-life case due to the increased salience of the relevant questions versus a guilty examinee in a laboratory scenario, would not hold true in at least those instances when a DLCQ was used. This is because the guilty examinee in a real-life case would associate the DLCQ as the means by which his or her physiological lie pattern is acquired for comparison with the relevant questions (Matte 1998; Matte & Reuss, 1999). Thus the response intensity of the DLCQ becomes related to the response intensity of the relevant questions, which in a real-life examination is significantly greater than in the laboratory. Hence the discussion of the DLCQ between tests has the effect of reinforcing the importance of the DLCQ and its ability to identify the guilty examinee's lie pattern for comparison to the relevant questions. The real-life fear of detection to the relevant question(s) can be transferred to the DLCQ, which offers the guilty examinee an equal if not greater threat of lie identification. Guilty examinees in real-life examinations have significantly more incentive to use countermeasures on test questions than guilty examinees in a mock crime paradigm. Recent survey research (Matte & Reuss, 1999) revealed that 86% of guilty participants considered the DLCQ an equal or greater threat than the relevant questions.

The use by Honts of several laboratory studies involving diverse polygraph techniques and methodology completely ignores that there may be many other significant factors responsible for the difference in accuracy and percentage of false negatives and positives. These factors could include the type of test used (multiple- or single-issue), types of comparison questions (current exclusive, non-current exclusive, non-exclusive, disguised, relevant-connected, directed-lie), the polygraph testing methodology including the pretest interview format, the test data analysis, and the competency of the polygraphist, to name a few. Furthermore, Honts' selection of studies where comparison questions were not discussed between charts is very limited and selective. The Szucko and Kleinmuntz (1981) study selected by Honts reflects the poverty of his selection process. The Szucko et al study used four examiner-trainees, which of itself should have eliminated the study from consideration inasmuch as it does not replicate a real-life examination. Furthermore, the integrity of the Szucko study has been seriously questioned, in that one of the participating examiners (Chodkowski, 1986) challenged the facts as they were published. It is interesting to note that nowhere in the Szucko study does it state that the comparison questions were not discussed with the examinees between charts. Yet Honts states in his Study Selection that "The studies shown in Table 1 were selected for inclusion in the analysis because they met at least one of the following criterion: The method section of the study explicitly described the discussion of, or the lack of discussion of, comparison and/or relevant questions between question list repetitions." Thus many other studies which reflected significantly higher accuracy rates could have qualified for inclusion in Honts instant article which at best can only be classified as selective scholarship.

In footnote #3, Honts states that "When I attended the Backster School of Lie Detection

in San Diego in 1976 the review of questions and the stimulation of comparison questions between charts was considered to be a standard practice." Honts previously made a similar statement under oath in U.S. v. Gilliard, to wit: "When I went through polygraph school at Backster, that was a standard part of the practice, to discuss the control (comparison) questions between the tests." In that latter statement, Honts mentions the discussion of the comparison questions only. In an e-mail to Honts, Dr. Ronald M. Reuss (1998) advised him that his testimony regarding the Backster school practice of discussing the comparison questions between tests was in error, based on a letter he had read from Backster (1998). Honts now has added the relevant questions in his description of the discussion that takes place between charts, but it should be noted that Honts' statement "the review of questions and the stimulation of comparison questions") confirms the emphasis is squarely on the comparison questions. Backster's letter, reprinted verbatim below, contradicts Honts' statements in U.S. v. Gilliard, and in his 1999 study:

> "After formulation and discussion of control (contrast) questions during the pre-test interview, further routine discussion of these questions will be avoided except as dictated by principles outlined in our Zone comparison Indication-Remedy table.
>
> 1. Should the examinee be reacting to relevant questions only (Combination A), there is no need for further stimulation on the control (contrast) questions through between-charts discussion. The adequacy of these questions can be verified by obtaining additional admissions following the last chart collected on that same target issue and prior to seeking target issue admissions due to the examinee's reactions to the relevant questions.
>
> 2. Should the examinee be reacting to the control (contrast) questions only (Combination B) there is no need for further "between charts" discussion of these questions.
>
> 3. Should the examinee be reacting to both the relevant questions and the control (contrast) questions (Combination D), further direct discussion of these questions could be counter-productive. In a more subtle fashion the examiner should reduce the intensity of these questions through a more indirect approach.*
>
> 4. Should the examinee show no reaction to any of the questions (Combination H) it would then be proper to attempt to stimulate reaction to the control (contrast) questions by directly discussing these questions between charts.
>
> As indicated on the enclosed "Tri-Zone" Reaction Combinations table, the above procedures have been a stable and consistent part of our polygraph examiner training material since 1962." (See attached "Tri-Zone" Reaction Combinations)
>
> *Note: This indirect approach requires that the examiner first review an irrelevant (neutral) test question with the examinee and make a cosmetic change. The examiner then reviews the relevant test questions with the examinee, which is followed by the review of the changed control questions. The focus is thus not on the control questions which are in some fashion changed to reduce their intensity. (Remedy listed in Combination D).

In summation, Honts acknowledges that "[c]orrelational studies and analyses are not as good as experiments in determining causation." The recently completed survey by Matte and Reuss (1999) suggests that the direct review or discussion of comparison questions between charts may increase the guilty examinee's apprehension regarding the DLCQ, thus creating a formula for false negative results.

References

Abrams, S. (1991). The directed-lie control question. *Polygraph*, 20(1), 26-31.

Abrams, S. (1999). A response to Honts on the issue of the discussion of questions between charts. *Polygraph*, 28(3), 223-227.

Backster, C. (1990). Backster Zone Comparison Technique chart analysis rules. Handout at lecture of the 25th annual seminar/workshop of the American Polygraph Association, 14 August 1990, Louisville, Kentucky.

Backster, C. (1998, September 18). Personal communication with J. A. Matte.

Chodkowski, R. E. (1986, February 22). Personal communication with Dr. Frank Horvath.

Honts, C. R., Gordon, A. (1999). A critical analysis of Matte's analysis of the directed lie. *Polygraph*, 27(4), 241-252.

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.

Honts, C. (1999). Discussion of Comparison Questions. *Polygraph*, 28(2), 117-123.

Horowitz, S. W., Kircher, J. C., Honts, C. T., Raskin, C. D. (1997). The role of comparison questions in physiological detection of deception. *Psychophysiology*, 34(1), 108-115.

Matte, J. A. (1998). An analysis of the psychodynamics of the directed-lie control question in the control question technique. *Polygraph*, 27(1), 56-67.

Matte, J. A., Reuss, R. M. (1999). Validation of potential response elements in the directed-lie control question. *Polygraph*, 28(2), 124-142.

Szucko, J. J., Kleinmuntz, B. (1981). Statistical versus clinical lie detection. *American Psychologist*, 36(5), 488-496.

"TRI-ZONE" REACTION COMBINATIONS

COMBINATION				INDICATION (Backster Zone Comparison Test)	REMEDY
A	(r)	dd	r	A1 PRESENCE OF RESPONSE TO ONE OR BOTH RED ZONE QUESTIONS INDICATES DECEPTION REGARDING TARGET ISSUE	A2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED
			g	A3 LACK OF RESPONSE TO BOTH GREEN ZONE QUESTIONS BECAUSE OF DAMPENING BY RED ZONE QUESTION RESPONSES INDICATES DECEPTION REGARDING TARGET ISSUE	A4 NO REMEDY NECESSARY; NO REASON TO BELIEVE GREEN ZONE QUESTION STRUCTURE INADEQUATE; GREEN ZONE QUESTIONS FUNCTIONING AS DESIGNED
			b	A5 LACK OF RESPONSE TO BOTH BLACK ZONE QUESTIONS INDICATES THAT NO OUTSIDE ISSUE BOTHERING SUBJECT	A6 NO REMEDY NECESSARY; EXAMINER HAS SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUE
B		tt	r	B1 LACK OF RESPONSE TO BOTH RED ZONE QUESTIONS INDICATES TRUTHFULNESS REGARDING TARGET ISSUE	B2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED
	(g)		g	B3 PRESENCE OF RESPONSE TO ONE OR BOTH GREEN ZONE QUESTIONS INDICATES TRUTHFULNESS REGARDING TARGET ISSUE, AS NO OTHER ZONE IS DAMPENING OUT GREEN ZONE	B4 NO REMEDY NECESSARY; NO REASON TO BELIEVE GREEN ZONE QUESTION STRUCTURE INADEQUATE; GREEN ZONE QUESTIONS FUNCTIONING AS DESIGNED
			b	B5 LACK OF RESPONSE TO BOTH BLACK ZONE QUESTIONS INDICATES THAT NO OUTSIDE ISSUE BOTHERING SUBJECT	B6 NO REMEDY NECESSARY; EXAMINER HAS SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUE
C		?	r	C1 LACK OF RESPONSE TO BOTH RED ZONE QUESTIONS USUALLY INDICATES TRUTHFULNESS REGARDING TARGET ISSUE; THIS RULE NULLIFIED BY BLACK ZONE QUESTION RESPONSE	C2 NO REMEDY NECESSARY; RED ZONE QUESTIONS WILL BE FUNCTIONING AS DESIGNED AFTER BLACK ZONE QUESTION RESPONSE SUBSIDES
			g	C3 LACK OF RESPONSE TO GREEN ZONE AND TO RED ZONE USUALLY INDICATES SERIOUS GREEN ZONE DEFECT. THIS IS NULLIFIED BY BLACK ZONE RESPONSE	C4 NO REMEDY NECESSARY; NO CAUSE TO BELIEVE GREEN ZONE QUESTION STRUCTURE INADEQUATE; RECHECK AFTER RESPONSE TO BLACK ZONE QUESTION SUBSIDES
	(b)		b	C5 PRESENCE OF RESPONSE TO ONE OR BOTH BLACK ZONE QUESTIONS INDICATES OUTSIDE ISSUE BOTHERING SUBJECT	C6 EXAMINER MUST GAIN SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUE
D	(r)		r	D1 PRESENCE OF RESPONSE TO ONE OR BOTH RED ZONE QUESTIONS INDICATES DECEPTION REGARDING TARGET ISSUE	D2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED
	(g)	?d	g	D3 PRESENCE OF RESPONSE TO ONE OR BOTH GREEN ZONE QUESTIONS IN ADDITION TO RED ZONE QUESTION INDICATES SERIOUS GREEN ZONE QUESTION DEFECT	D4 REDUCE INTENSITY OF GREEN ZONE QUESTIONS BY ALTERING SUBJECT AGE CATEGORIES OR CHANGING SCOPE OF GREEN ZONE QUESTIONS
			b	D5 LACK OF RESPONSE TO BOTH BLACK ZONE QUESTIONS INDICATES NO OUTSIDE ISSUE BOTHERING SUBJECT	D6 NO REMEDY NECESSARY; EXAMINER HAS SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUE
E	(r)		r	E1 PRESENCE OF RESPONSE TO ONE OR BOTH RED ZONE QUESTIONS INDICATES DECEPTION REGARDING TARGET ISSUE	E2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED
		d	g	E3 LACK OF RESPONSE TO BOTH GREEN ZONE QUESTIONS BECAUSE OF DAMPENING BY RED ZONE QUESTION RESPONSE; INDICATES DECEPTION REGARDING TARGET ISSUE	E4 NO REMEDY NECESSARY; NO REASON TO BELIEVE GREEN ZONE QUESTION STRUCTURE INADEQUATE; GREEN ZONE QUESTIONS FUNCTIONING AS DESIGNED
	(b)		b	E5 PRESENCE OF RESPONSE TO ONE OR BOTH BLACK ZONE QUESTIONS INDICATES OUTSIDE ISSUE BOTHERING SUBJECT	E6 EXAMINER MUST GAIN SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUES
F	(r)		r	F1 PRESENCE OF RESPONSE TO ONE OR BOTH RED ZONE QUESTIONS INDICATES DECEPTION REGARDING TARGET ISSUE	F2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED
	(g)	?d	g	F3 PRESENCE OF RESPONSE TO ONE OR BOTH GREEN ZONE QUESTIONS IN ADDITION TO RED ZONE RESPONSE INDICATES SERIOUS QUESTION DEFECT IN GREEN ZONE QUESTIONS	F4 REDUCE INTENSITY OF GREEN ZONE QUESTIONS BY ALTERING AGE CATEGORIES OR CHANGING SCOPE OF GREEN ZONE QUESTIONS
	(b)		b	F5 PRESENCE OF RESPONSE TO ONE OR BOTH BLACK ZONE QUESTIONS INDICATES OUTSIDE ISSUE BOTHERING SUBJECT	F6 EXAMINER MUST GAIN SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUE
G			r	G1 LACK OF RESPONSE TO BOTH RED ZONE QUESTIONS INDICATES TRUTHFULNESS REGARDING TARGET ISSUE	G2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED
	(g)	t	g	G3 PRESENCE OF RESPONSE TO ONE OR BOTH GREEN ZONE QUESTIONS INDICATES TRUTHFULNESS REGARDING TARGET ISSUE; NO OTHER ZONE IS DAMPENING OUT GREEN ZONE	G4 NO REMEDY NECESSARY; NO CAUSE TO BELIEVE GREEN ZONE QUESTION STRUCTURE INADEQUATE; GREEN ZONE QUESTIONS FUNCTIONING AS DESIGNED
	(b)		b	G5 PRESENCE OF RESPONSE TO ONE OR BOTH BLACK ZONE QUESTIONS INDICATES OUTSIDE ISSUE BOTHERING SUBJECT	G6 EXAMINER MUST GAIN SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUE
H			r	H1 LACK OF RESPONSE TO BOTH RED ZONE QUESTIONS STILL INDICATES TRUTH REGARDING TARGET ISSUE; THIS SYSTEM BASED ON SUBJECT CAPABILITY OF RESPONSE	H2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED
		t?	g	H3 LACK OF RESPONSE TO BOTH GREEN ZONE QUESTIONS IN ADDITION TO LACK OF RESPONSE TO RED ZONE QUESTIONS INDICATES SERIOUS GREEN ZONE QUESTION DEFECT	H4 INCREASE INTENSITY OF GREEN ZONE QUESTIONS BY ALTERING AGE CATEGORIES OR CHANGING SCOPE OF GREEN ZONE QUESTIONS
			b	H5 LACK OF RESPONSE TO BOTH BLACK ZONE QUESTIONS INDICATES NO OUTSIDE ISSUE BOTHERING SUBJECT	H6 NO REMEDY NECESSARY; EXAMINER HAS SUBJECT CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUE

Can We Trust Counterintelligence Polygraph Tests?

Vance MacLaren

Abstract

Polygraph interviews are an important part of security clearance procedures in many branches of the American government. This paper briefly reviews open source literature on polygraph security screening procedures currently in use by the US Department of Defense. Current polygraph security screening practices make a valuable contribution to the maintenance of national security.

Key words: counterintelligence screening, TES, Test for Espionage and Sabotage, validity.

Recent developments at America's national laboratories have drawn public attention to the issue of polygraph security screening. In the wake of allegations of penetration of the nuclear lab's security by the People's Republic of China, the Department of Energy has initiated polygraph counter-intelligence screening of some employees with access to sensitive information. Some critics suggest that such tests are inaccurate, and that they infringe upon the dignity of government employees (see http://www.stop polygraph.com online for criticisms of security screening polygraph tests). In the present paper, it is argued that current polygraph security screening practices make a valuable contribution to the maintenance of national security.

Recent History

Polygraph interviews have been an integral part of security clearance procedures for both civilian and military personnel since the early days of the Cold War. They are also widely used in pre-employment screening of police recruits (Meesig & Horvath, 1995). Fear of an increase in espionage activity aimed against the United States led to a healthy skepticism of security vetting procedures in the 1980s (e.g. United States Congress Office of Technology Assessment, 1983). Polygraph screening techniques in use at that time were subjected to systematic appraisal in which several studies (e.g. Barland, Honts & Barger, 1989; Honts, 1992) found the Counter-intelligence Scope Polygraph (CSP) technique to be inadequate as a means of identifying persons involved in espionage activity. As the magnitude of the Aldrich Ames spy case was revealed in the early 1990s, it became clear that something had to be done to improve the methods used to detect subversives operating within the government and the military.

Around 1992, the Department of Defense Polygraph Institute (DoDPI) began development of an improved test to be used in security screening (DoDPI, 1994). The result was the Test for Espionage and Sabotage (TES). The TES format differs in many ways from the CSP, and is generally more standardized and less intrusive than older methods. Also, it makes use of Directed-lie comparison questions (Abrams, 1991; Honts & Raskin, 1988; Horowitz, Kircher, Honts & Raskin, 1997). According to the DoDPI annual report to Congress for fiscal year 1994 (DoDPI, 1995), by that time, "All DoD agencies [had] switched from the CSP to TES for security screening" (p. 28).

Laboratory Validity

To date, only two large-scale studies designed to evaluate the accuracy of TES have been published (DoDPI Research Division Staff, 1997; 1998; see also Reed, 1994). In

Author Note

Special thanks to George Maschke and Dr. Charles Honts. Requests for reprints may be sent to Vance MacLaren, Department of Psychology, University of New Brunswick, P.O. Box 5050, Saint John, N.B., Canada, E2L-4L5, or by electronic mail to vancem@nbnet.nb.ca.

both studies, participants in programmed guilty groups were required to take part in very realistic simulations of espionage activity. In the two studies, 50 of the 60 guilty participants were correctly identified (83.3%). Of the 108 participants in programmed innocent groups, 98 were correctly identified (90.7%). Although further replication of these findings should be sought, it appears that the TES procedure has an acceptable level of discriminative accuracy under controlled laboratory conditions.

Field Validity

At present, no systematic field study of TES has been published, but some information is available in the form of DoDPI's annual reports to Congress, which present statistics on the activities of the Department of Defense screening program. In the last five fiscal years, the Department of Defense has administered 43648 counterintelligence screening tests. This does not include testing performed under authority of other federal agencies, such as the National Security Agency and National Reconnaissance Office. In those tests, 902 people made admissions relevant to security issues. In addition, 148 of the tests produced either "significant physiological responses" or "inconclusive" outcomes. Of the 963 cases in which either relevant admissions were made, or in which inconclusive or significant physiological responses were found, only 24 resulted in adverse action following subsequent investigation. An additional 97 were still pending investigation or adjudication at the time the reports were presented to Congress, so it is not known how those cases were resolved. The actual number of cases in which investigation subsequent to the polygraph screening revealed evidence sufficient to result in adverse action can therefore be estimated as falling between 24 ad 121.

The tiny percentage (less than 0.3%) of cases in which access to information was denied or withheld may seem like a meager harvest, in light of the expensive resources devoted to the polygraph screening program. However, if one considers the scale of damage that can be caused by individuals like Edwin Pitts, David Boone, or Harold Nicholson, even these small numbers are anything but insignificant (see http://intellit.muskingum.edu/online for information on these and other spy cases). It should also be remembered that many of the individuals identified by the polygraph screening program were also subjected to background investigations and other means of screening, which failed to identify their illegal activity. In addition, others working for the government may have been deterred from engaging in espionage activity in the first place, out of fear of being detected by periodic or aperiodic screening.

The False Negative Problem

In a CSP study by Barland, Honts & Barger (1989), approximately 20% of the 207 participants made admissions about real-world violations of security protocols in the course of their mock-espionage CSP examinations. Based on this estimate, Honts (1994) erroneously concluded that the CSP program had an extraordinarily large rate of false negative error. While a large proportion of employees may violate security procedures at some time in their careers, the vast majority of such violations are of trivial importance. They may indicate sloppy security practices, but certainly not involvement with organized espionage. It is laughable to suppose that one in five government employees is a spy. A more reasonable estimate of the prevalence of espionage that one might draw from the Barland et al. study is the number of security violations considered "significant". After all, the purpose of the screening program is not to detect petty infractions of institutional policy; it is to protect national security from legitimate threats. Of the 207 persons examined in that study, only 2 made such admissions (1%). This rate is similar to that found in the Department of Defense screening program, in which admissions were obtained in 902 of 43648 cases (2%).

If we follow the example of Honts' conditional probability analysis and assume that 1% of persons screened by the polygraph are actually guilty of taking part in activities that might undermine national security, we might then assume that about 436 of the 43648 people tested under the Department of Defense polygraph screening program had committed security violations of some consequence. Because only between 24 and

121 were actually identified, according to the DoDPI report, it seems reasonable to concur with Honts' assertion that a large number of serious security violations may remain undetected, although the prevalence of false negatives is probably much less severe than he had supposed.

In any case, most of those who received adverse action subsequent to their polygraph interview would not have been detected if the screening program were not in place. No one technology can be expected to detect all subversives, but if each of the security measures used by the federal government (e.g. financial records analysis, background interviews, etc.) manages to identify a few, then they may collectively provide an effective obstacle against those who would do harm to national security.

Two Strikes and You're Out... Maybe

In the two simulation studies of TES, 90.7% of innocent participants were correctly identified. If we tentatively assume that these results generalize to field conditions, then we must conclude that 9.3% of innocent government employees would be falsely accused of espionage by a TES exam. However, that does not appear to have happened, according to the statistics reported in the DoDPI annual reports to Congress. Why, then, are so few people classified as deceptive or inconclusive by TES? Without a proper field study of the TES procedure, it is not correct to rule out the possibility that the detection rates found in the simulation studies might not generalize to the field situation. Another possibility, as suggested by Barland, Honts & Barger (1989), is that some important precautions are used to protect employees from false accusation. No test is immune to error but, if the test is applied carefully, the negative ramifications of errors may be minimized.

One simple way to minimize possible harm to employees would be to require repeated examinations whenever significant responses or inconclusive results are found. With repeated administrations of a test, the likelihood of being falsely classified by each repetition tends to diminish. Suppose that 10,000 innocent people are given TES examinations and that 930 (9.3%) fail the test. To preclude the possibility of falsely smearing those individuals' careers, all 930 are re-tested. This time, 844 (90.7%) of them pass the test, and 86 (9.3%) fail it. The prior likelihood that an individual would fail either of two tests is p = .093 +.093 = .186, but the likelihood that an individual would fail both of the tests is only p = .093 * .093 = .0086. By simply enacting the policy that all individuals who fail one test are subsequently re-tested to verify the results, the likelihood of false positive error is greatly reduced. Under this scheme, over 99% of innocent employees would pass the security screening. Whatever the true specificity of TES, and whatever the effect of prior testing on the likelihood of false positive error in subsequent interviews, a requirement of corroborative results from repeat testing should tend to ameliorate the chances of falsely accusing innocent employees. One important, and as yet unanswered question, is the level of test-retest reliability of results found in the TES screening procedure.

By the same logic, we would expect 83.3% of guilty operatives to be identified by the first test. These employees would be subjected to added scrutiny, perhaps including the initiation of other investigative procedures, any of which might reveal evidence of their illegal activity. In any case, the chance that a spy would fail both tests is p = .833 * .833 = .694, or approximately 69%.

Now suppose that we were to randomly test 10,000 employees, and for simplicity sake, 1% of them are spies. On the first test, 921 of the 9,900 innocent employees and 83 of the 100 guilty ones fail. On a second test, 86 innocent people fail the test, and 69 guilty ones fail. We now have 155 cases of people who failed the test twice. Of these, 45% are foreign agents, and 55% are not. Additional investigative techniques would need to be deployed to sort the innocent from the guilty. However, the investigators now have only 125 cases to sift through; not 10,000. In this scenario, use of the polygraph produces tangible benefits to the efficiency and cost-effectiveness of counterintelligence efforts.

By applying the counterintelligence polygraph tests in an orderly and careful way,

large numbers of individuals can be effectively screened. This must lead to reductions in the financial expenditures related to security investigations. However, care must be exercised because false negative errors can occur, and if a person passes a polygraph screening, they should not be exempted from other screening procedures.

Admissions

There is an important distinction that must be made between criminal investigative applications of polygraph tests and the types of situations involved in security screening. Security screening tests are not specific-issue tests. They are a way to probe large numbers of employees in search of a tiny proportion that are involved in undesirable security breaches. It is a needle-in-a-haystack problem. In some respects, the accuracy of the technique may be of importance secondary to its utility. Whereas the identification of individual saboteurs and spies may be important, knowing the nature of their activities may be equally useful. Admissions obtained in the course of polygraph testing may be a valuable source of counterintelligence information. Consider the following excerpt from the Office of Technology Assessment report (1983):

> "It appears that NSA (and possibly CIA) use the polygraph not to determine deception or truthfulness per se, but as a technique of interrogation to encourage admissions. NSA has stated that the agency does not use the 'truth v. innocence' concept of polygraph examinations commonly used in criminal cases. Rather, the polygraph examination results that are most important to NSA security adjudicators are the data provided by the individual during the pretest or posttest phase of the examination." (p.100).

In some cases, such "data" may be very important, indeed. Whether or not the squiggly lines produced by a polygraph can detect deception is an academic dispute that bears little importance to the applied issue of whether the technique is a useful, fair and cost-effective contributor to protecting the nation's most treasured secrets.

Conclusions

The polygraph technology used in the Department of Defense screening program appears to have considerable validity and utility. Previous assessments of the situation (e.g. Honts, 1994; Honts, 1991) were written at a time when older techniques (i.e. CSP) were in use. The newer TES procedure appears to be a great improvement over those methods. The results of actual examinations summarized in the DoDPI annual reports to Congress are not consistent with the idea that large numbers of innocent public servants, civilian contractors and military personnel are being unfairly "fluttered" out of their jobs (e.g. Lykken, 1998). The polygraph screening program has a respectable record of identifying persons with hostile or selfish intentions that are contrary to the maintenance of national security.

References

Abrams, S. (1991). The directed lie control question. *Polygraph*, 20, 26-31.

Barland, G.H., Honts, C.R. & Barger, S.D. (1989). Studies of the Accuracy of Security Screening Polygraph Examinations. Department of Defense Polygraph Institute, Fort McClellan, AL.

Department of Defense Polygraph Institute (1994). Polygraph program, United States Department of Defense. *Polygraph*, 23, 61-84.

Department of Defense Polygraph Institute (1995). Department of Defense polygraph program report to congress for fiscal year 1994. *Polygraph*, 24, 13-33.

Department of Defense Polygraph Institute (1996). Fiscal year 1995 report to Congress on the Department of Defense polygraph program. Ft. Jackson, South Carolina: Author.

Department of Defense Polygraph Institute (1997). Department of Defense polygraph program annual report to congress for fiscal year 1996. *Polygraph*, 26, 240-254.

Department of Defense Polygraph Institute Research Division Staff (1997). A comparison of detection accuracy rates obtained using the counterintelligence scope polygraph and the test for espionage and sabotage question formats. *Polygraph*, 26, 79-106.

Department of Defense Polygraph Institute (1998). Fiscal year 1997 report to Congress on the Department of Defense polygraph program. *Polygraph*, 27, 171-180.

Department of Defense Polygraph Institute Research Division Staff (1998). Psychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage. *Polygraph*, 27, 68-73.

Department of Defense Polygraph Institute (1999). Fiscal year 1998 report to Congress on the Department of Defense polygraph program. Ft. Jackson, South Carolina: Author.

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. (1992). Counterintelligence scope polygraph (CSP) test found to be poor discriminator. *Forensic Reports*, 5, 215-218.

Honts, C.R. (1988). A field study of the validity of the directed lie control question. *Journal of Police Science and Administration*, 16, 56-61.

Honts, C.R. (1994). Psychophysiological detection of deception. *Current Directions in Psychological Science*, 3, 77-82.

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.

Lykken, D.T. (1998). *A Tremor in the Blood: Uses and Abuses of the Lie Detector*. New York: Plenum.

Meesig, 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, 57-136.

Reed, S. (1994). A new psychophysiological detection of deception examination for security screening. *Psychophysiology* (Abstract), 31, S80.

United States Congress Office of Technology Assessment (1983). Scientific Validity of Polygraph Testing: A Research Review and Evaluation - A Technical Memorandum, OTA-TM-H-15. Washington, D.C.: U.S. Government Printing Office.

The Hybrid Directed-Lie Test, The Overemphasized Comparison Question, Chimeras and Other Inventions: A Rejoinder to Abrams (1999)

Charles R. Honts, David C. Raskin, Susan L. Amato, Anne Gordon, & Mary Devitt

Abstract

Abrams (1999) claims to be a response to Honts (1999) on the matter of the discussion of questions between repetitions in comparison question tests. However, Abrams fails to even mention the data from the 1,092 polygraph examinations reviewed by Honts (1999). Instead, Abrams (1999) uses a term for something he calls the hybrid directed-lie test, but fails to provide an operational definition of that test. Abrams (1999) then uses largely inaccurate anecdotes to try to prove that the directed-lie chimera he created is an inaccurate test. Abrams' (1999) descriptions of research and case facts are frequently misleading and often inaccurate. The present paper corrects the record by providing accurate descriptions of the scientific research and the cases that were misrepresented by Abrams (1999). Nothing in Abrams' article questions the validity of the previous scientific research and conclusions reported in Honts (1999) or Honts and Gordon (1998).

In an (1999) article in this journal, Abrams published an article regarding the discussion of questions between question repetitions in psychophysiological detection of deception examinations of the comparison question type. However, instead of providing a discussion of the scientific data concerning this issue, Abrams (1999) engages in an ad hominem attack on the methods used in the private practices of Drs. Raskin and Honts. Moreover, Abrams (1999) focuses his attack more on the directed-lie comparison (DLC) question than on the discussion of questions between repetitions. The nature of this attack is not scientific, but rather consists of a list of anecdotes, which Abrams claims support his position, while in reality they do not. Finally, Abrams (1999) invective is full of inaccuracies and misrepresentations.

The present article is being published to correct the record, and is divided into four sections. In the first section we note Abrams' (1999) arguments against the review of questions between question repetitions. Since science is based on data and not on personal belief, nor on calls to authority, we revisit the data on the discussion of questions between repetitions, data which Abrams (1999) ignored. In the second section, since Abrams (1999) is more focused on the DLC than on the review of questions between repetitions, we also revisit the validity data on the DLC. Moreover, we correct the misrepresentations Abrams (1999) makes about those studies. In the third section we address the actual cases that Abrams (1999) cites as his evidence that the Raskin and Honts field practice methods are invalid. Here we also spend considerable time correcting the record for Abrams' (1999) extensive errors and misrepresentations. In the final section we summarize the science and point out the fatal flaws in Abrams (1999) polemic.

1. The Science Concerning the Review of Questions Between Question Repetitions

Abrams (1999) presents no new data regarding the effects of the review of questions. Instead he states his personal belief that:

Charles R. Honts, Ph. D., Boise State University; David C. Raskin, Ph. D., University of Utah; Susan L. Amato, Ph. D., Anne Gordon, Ph. D., Boise State University; Mary Devitt, Ph. D., University of Minnesota. Send reprint requests to Dr. Honts, Psychology Dept, Boise State University, 1910 University Dr., Boise, ID 83725. E-mail to Honts@truth.idbsu.edu.

> In essence all of this means that there is a delicate balance that exists between the comparison and relevant questions and many variables can tip this balance in either of those two directions. Too much discussion of one or the other during the pretest, a difference in inflection or loudness when the questions are being asked, any discussion between the charts that stresses either the relevant or comparison questions, or any mental activity on one questions versus another can weigh the balance in the direction of that particular emphasis. (p. 224)

Given the extreme frailty that Abrams attributes to comparison questions tests in, what we will refer to as his Unbalanced Hypothesis (UH), one is left wondering how such a test could ever be administered in a valid manner. The UH sounds much more akin to the criticisms of Furedy and Ben-Shakhar (1991) rather than critical commentary from someone who is an advocate of the polygraph profession. Fortunately for the polygraph profession, Dr. Abrams' (1999) assessment of the frailty of the CQT is absolutely without empirical support.

Honts (1999) reviewed the existing data on the discussion of questions between repetitions. He found 19 studies of comparison question tests[1] where the discussion of questions between charts could be determined. In eight of the studies no discussions occurred between question repetitions, in the other 11 studies questions were discussed between the question repetitions (see Honts, 1999, Table 1, p. 123). These studies included 1092 polygraph examinations, and the data from those 1092 examinations profoundly contradict Abrams' (1999) UH. Significantly greater accuracy for both innocent and guilty subjects was associated with a review of questions between questions.

Abrams (1999) could have argued, but did not, that a balanced review of questions between charts would not be a problem. However, there were two studies (Dawson, 1981; Patrick & Iacono, 1989) included in the Honts (1999) analysis that directly addressed Abrams' (1999) UH that the unequal emphasis of comparison over relevant questions will result in an increase in false negative outcomes. In the Dawson (1981) and the Patrick and Iacono (1989) studies the examiners discussed the comparison questions between the question repetitions but explicitly did not discuss the relevant questions. The data from those two studies are reproduced here in Table 1. If Abrams' (1999) hypothesis about the frailty of the comparison question test to unbalanced discussion was correct, we would expect that the false negative rates in the Dawson (1981) and Patrick and Iacono (1989) studies would be extremely high. In Dawson (1981) there

Table 1. Two Studies Where Comparison But Not Relevant Questions Were Discussed Between Repetitions.

	Guilty				Innocent			
Study	n	% Cor	% Wrong	% Inc	n	% Cor	% Wrong	% Inc
Dawson (1981)	12	100	0	0	12	75	17	8
Patrick and Iacono (1989)	24	83	13	4	24	33	42	25
Total n and unweighted means	36	91.5	7.5	2.0	36	54.0	29.5	16.5

[1] Honts (1999) included studies of both probable-lie and directed-lie comparison questions. According to Abrams (1999) the type of comparison question should not matter. According to Abrams' (1999) UH, it is the alleged overemphasis on the comparison questions in the highly frail comparison question test that is the danger.

were no false negative errors and no inconclusive outcomes with guilty subjects. This high performance occurred despite the prediction that follows from Abrams' (1999) UH that the methods used by Dawson should have produced an unacceptably high false negative rate. The overall performance in the Patrick and Iacono (1989) study was not very good, but in spite of Abrams (1999) predictions about the frailty of the comparison questions test, false positive errors greatly outnumbered false negative errors. Across the two studies, the false positive rate was almost four times higher than the false negative rate.

In science, progress is measured by testing hypotheses by attempting to find data that falsify them (show them to be incorrect). This is why the question of the existence of a testable hypothesis and whether the hypothesis has been tested are part of the Daubert (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 1993) standard for the admission of scientific evidence in U. S. Federal courts. Once data are found that falsify a hypothesis, the hypothesis has to be abandoned or modified to take the data into account. This is the basic process of all science. Abrams' (1999) UH (emphasizing the comparison questions over the relevant questions will result in a high rate of false negative errors) was directly tested in two scientific studies that were published in high quality scientific journals. The UH was clearly shown to be false. Abrams' (1999) UH should be abandoned because it has clearly been falsified by data.

Honts (1999) noted that during Abrams' testimony about the frailty of the comparison question test, he was unable to cite a single study that supported his position (see p. 118, Honts 1999). Abrams (1999) retorted that Honts failed to mention Abrams (1991) as an exemplar of such research. Moreover, Abrams (1999) claimed that the methods used in Abrams (1991) were representative of those used by Raskin in his field practice. Nothing could be further from the truth. Abrams (1991) was a field experiment with only 10 subjects (6 deceptive). Only one directed-lie question was asked only one time in the very last position on the last question repetition. The DLC presentation and manipulation were made just before the final repetition where the one and only DLC question was going to be asked. Abrams violated the generally accepted basic procedures and rules of question presentation for any comparison question test, directed-lie or probable-lie. To suggest that his procedures are representative of the methods used by Honts and Raskin either in their field practices or in the validity studies discussed below is disingenuous.

Honts (1999) did not discuss Abrams (1991) because the Abrams study did not address the review of comparison questions between repetitions. Moreover, in sworn testimony Abrams described his own 1991 study as follows:

> Q. Would you agree that your study represents too small a sample to make generalizations from?
> A. Worse than that. It's that the directed-lie is only -- only occurs one time at the end of the test, and that certainly weakens it, and it's indicated in that paper.
> Q. In fact, you only used ten subjects?
> A. That's correct.
> Q. And of those ten subjects, you only used the directed-lie on one of the three charts that you ran on each subject?
> A. That's correct.
> Q. And because of the very small amount of data, Professor Honts felt like that the generalizations you were making in your paper and which you have made today were not justified. Isn't that what he indicated?
> A. That's, that's true. I didn't read his whole critique. I just saw it for the first time. But that's certainly true...
> (Trial Transcript *U. S. v. Gilliard*, 1996).

Then in the case of *United States v. Walker* (1999) the following exchange took place between the cross-examining attorney and Dr. Abrams:

> Q. Does your study, your 1991 study, does it address talking between charts?
> A. Not that specifically.

When Abrams (1999), in referring to Abrams (1991), states, "Some of these changes, particularly with deceptive subjects,

were quite strong, indicating that placing emphasis on the HDLQ through a discussion between charts could certainly influence even deceptive subjects to appear more truthful . . ." (p. 225), he directly contradicts his own sworn testimony. Given that Abrams agrees that generalizing from his 1991 paper to the techniques used by Honts in Gilliard was unjustified, it now seems quite odd that he takes Honts (1999) to task for not citing his pilot DLC study in a discussion where that study, by Abrams' own description, wasn't relevant.

Misrepresentations Regarding the Review of Questions in Abrams (1999).

Abrams (1999) cites a warning in Fuse (1982) that if too much emphasis were placed on the directed-lie questions this could result in false negative outcomes. However, Abrams (1999) conveniently failed to mention that Fuse (1982) recommends the stimulation of the directed-lie questions between repetitions.

> If it appears that the DLCs are not generating at least some response activity, regardless of the response activity to the relevants mild interchart reinforcement may be utilized. For example, examinee may be told the test chart "looks good" and that when he lies, their responses are clear, and it is obvious that he has indeed engaged in the activities covered by the DLCs. (p. 25)

Despite Abrams' (1999) assertions to the contrary, Fuse (1982) actually suggested the need for discussion of the directed-lie questions between repetitions.

Abrams (1999) quotes a section from Honts and Perry (1991) that acknowledges the possibility that a dishonest and unethical examiner could manipulate an examination to deliberately produce a desired outcome. However, Abrams (1999) takes this comment out of context and misrepresents it in his article. The Honts and Perry statements were made in the context of an argument for the tape recording of all polygraph examinations. The paragraph that Abrams (1999) quotes ends with the following sentence, "These types of manipulations might be very difficult to uncover, unless some permanent record of the polygraph examination was made and offered for scrutiny." (Honts & Perry, 1991, p. 372) Two points are critical here. The first is that the efforts at manipulation referred to by Honts and Perry are deliberate and unethical acts. The second point is that an examiner engaged in such unethical behavior would attempt to hide his or her manipulations by not creating a permanent record of the examination by tape recording it. This is a very different situation from using a standardized test in the manner in which it has been validated in laboratory and field studies and preserving a complete record of the examination by tape recording. Drs. Honts and Raskin have always tape-recorded their polygraph examinations and they have always been willing to have those tapes reviewed under the appropriate legal circumstances. Moreover, they are strong public advocates for the taping of all polygraph examinations (Honts & Perry, 1991; Raskin, 1986). The tape recording of all polygraph examinations is not a practice that Dr. Abrams follows (Griffith v. Melgaard, 1995).

2. The Scientific Evidence Concerning the Validity of the Directed-Lie Comparison Question Test.

The available scientific research on the validity of the DLC was reviewed extensively in this journal by Honts and Gordon (1998). With the exception of the Abrams (1991) study, which Abrams himself describes as a flawed pilot study that is not generalizable, we know of no published research that has indicated the directed-lie comparison question produces results that are less accurate than those produced by tests that use probable-lie comparison questions.

Moreover, the U. S. Government has recently adopted a directed-lie test for its recent large expansion of national security screening to the Department of Energy (Beardsley, 1999). It should be noted that the main concern in a national security screening program must be the false negative rate. A false negative error in the national security screening program can have catastrophic costs, whereas a false negative error in the legal system is considered to be less serious than a false positive error. Clearly, the U. S. Government has decided that testing with a directed-lie test is the best way to protect the

national security system from false negative errors. This decision is supported by research that shows higher validity for the directed-lie as compared to probable-lie test (Reed, 1994; Department of Defense Polygraph Institute Research Division Staff, 1997 & 1998). Additional reviews and discussion of the validity research on the DLC can be found in Raskin, Honts, and Kircher (1997).

Misrepresentations of Directed-lie Research in Abrams (1999).

Abrams (1999) criticizes the Horowitz, Kircher, Honts, and Raskin (1997) study for producing what he describes as a low percentage of correct outcomes for guilty subjects (73%). However, percentage accuracy by itself is not a good measure of the discriminative power of a diagnostic test like a polygraph examination. A detection efficiency coefficient (Kircher, Horowitz, & Raskin, 1988) is a much better index because it quantifies the ability to discriminate the truthful from the deceptive and takes into account inconclusive rates. The detection efficiency coefficient for the probable-lie tests in Horowitz et al., (1997) was 0.56, while the coefficient for the personally relevant directed-lies (the type used by Dr. Raskin and Honts in their field practices) was 0.69. This is a statistically reliable difference in favor of the directed-lie. The detection efficiency co-efficient also makes it easy to compare the probable-lie performance in Horowitz et al. with other studies of the probable-lie comparison question test. Kircher et al., (1988) reviewed the data from 14 laboratory studies of the validity of the probable-lie comparison test and calculated detection efficiency coefficients for each. Those values ranged from 0.21 to 0.87, with an unweighted mean of 0.67 and a standard deviation of 0.18. Although, the probable-lie result in Horowitz et al is slightly below the mean found for other studies, it is not significantly so. Clearly the results of Horowitz et al., (1997) are within the range of sampling variability for the population of studies of the probable-lie comparison questions test.

Abrams (1999) suggests caution in generalizing the results of Horowitz et al., because " . . . research in the field usually demonstrates about 95% accuracy for deceptive subjects . . ." (p. 225). No authority was provided for the 95% figure. While we agree that high quality field studies of comparison question tests (including the directed-lie) produce high accuracy rates, there is a range in detection accuracy. In the population of field studies reviewed by the OTA (1983) the accuracy of correct detection of guilty subjects ranged from 70.6% to 98.6% (p. 52). Thus, Abrams (1999) criticism of the generalizability of the Horowitz et al., (1997) study lacks merit.

Abrams (1999) then states the following criticism of the Honts and Raskin (1998) field study, "Since 6 of the 13 truthful subjects were accused of the sexual abuse of children, and knowing how frequently children who were actually abused recant, this would place a considerable degree of doubt on their so called confirmed truthful subjects." (p. 225) Abrams should know by now that this criticism is inaccurate and highly misleading. As Honts and Gordon (1999) reported in this journal, Abrams' criticism does not apply to 4 of the 6 innocent subjects as the method of confirmation was based on something other than a simple recantation. Despite this, Abrams (1999) presents information in his critique of Honts and Raskin (1988) that he should now know is incorrect.

Abrams' (1999) criticism of Honts and Raskin (1988) contains a second serious misrepresentation of psychological science. Abrams (1999) states the following premise in his criticism, " . . . knowing how frequently children who were actually abused recant, . . . (p. 225). No attribution is made for this purported statement of fact, and psychological science provides no support for such a claim. The notion that abused children frequently recant is part of the now discredited Child Sexual Abuse Accommodation Syndrome (CSAAS; Summit, 1983). Even Summit now denies that CSAAS has diagnostic validity (1992). A consensus statement by an international group of recognized experts in the area of the investigation of child sexual abuse (Lamb, 1994) stated conclusively that there are no behavioral syndromes associated with child sexual abuse. In addition, recent research by Bradley and Wood (1996) has shown that the rate of recantation by child sexual abuse victims is only 4% in those cases where the child had made an accusation.

Bradley and Wood (1996) conclude that, "The Child Sexual Abuse Accommodation syndrome described by Summit (1983) seems to be infrequent among the types of cases seen by child protection agencies." (p. 881) There simply is no scientific evidence to support Abrams' (1999) contention that child sexual abuse victims recant frequently or that such recantations are frequently false when they occur.

Our position is not that all child sexual abuse recantations are true. The data indicate that under infrequent circumstances they may be false. Similarly, suspects of crimes sometimes give false confessions under interrogation. No scientific evidence demonstrates that child sexual abuse recantations are any more suspect than are interrogation elicited confessions. Both can be false but most of the time they are valid. One of the major problems with field studies is that the criteria for confirmation may be weak (see Honts, 1996 for a validation of the use of information other than confessions in conducting field validity studies of the polygraph). However, if we are to abandon recantations as a criterion, then we will have to also abandon confessions, because they are equally suspect.

Abrams quotes Dollins (1998) in an effort to show that the DODPI does not believe that the validity of the DLC has been established. However, Dollins (1998) is not a review of the literature or a policy statement, rather it was an indication of areas where DoDPI was interest in funding research.

3. The Cases

Abrams (1999) cites a number of court cases that he says indicate problems with the DLC. Abrams (1999) suggests that because individuals may have taken plea bargains or have been eventually convicted of a crime, this is evidence for the accuracy or inaccuracy of any polygraph examinations conducted in that case. Abrams' (1999) position is universally rejected within the scientific community. Case outcomes, like plea bargains and verdicts may occur for many reasons. Innocent people do sometimes plea guilty to avoid more serious charges and penalties. In addition, polygraph examiners who work for defense counsel are often called upon to conduct tests on some aspect of a case that tests the counsel's theory of the case. It is often possible that the subject is tested, passes, the result is accurate, the trier of fact accepts the polygraph outcome as valid, and still convicts the defendant because of the requirements of the law. For all of these reasons, case outcome is never used in scientific research as a criterion of confirmation for polygraph examination accuracy.

Abrams' (1999) argument by anecdote is further flawed because his descriptions of these cases are often grossly inaccurate and misleading. Although, argument by anecdote is not a scientific approach[2] and is not useful for determining the validity of any polygraph technique, we believe that it is necessary to set the record straight on each of these cases so the readership of *Polygraph* will have access to accurate information. We address the cases in the order Abrams (1999) cited them.

U. S. v. Gilliard (1996). Dr. Honts conducted a test on the defendant in a Medicare/Medicaid fraud case. The test used contained both directed-lie and probable-lie comparison questions. Abrams' (1999) description of this case contains misrepresentations. Abrams stated that the reason the polygraph was not admitted in Gilliard was that the research on the DLC was minimal and conflicting. However, Abrams (1999) failed to mention that Gilliard had also taken another polygraph examination that was administered by a law enforcement officer who used the probable-lie technique. Defendant Gilliard passed both polygraph examinations. Abrams (1999) also failed to mention that the Magistrate who held the Daubert hearing concluded the following regarding the DLC:

[2] Even if we wanted to argue our position by anecdote, we are at a serious disadvantage. Raskin and Honts could cite many confirmed high profile cases where subjects were tested and produced strong deceptive results with the DLC. However, since most of their private practice work is for defense attorneys they are ethically restricted from naming cases where the subjects have failed and confessed.

> Based on the current record, therefore, the Court accepts the testimony of Dr. Honts, which is based on empirical data generated by a published study subjected to peer review, that the error rate for the hybrid approach is approximately the same (if not better than) the error rate for the probable-lie technique, which is approximately ten percent. (Smith, 1996, p.34).

The Magistrate then issued the following order, "For the reasons discussed herein, the Court GRANTS the government's motion to exclude the results of the polygraph examination administered by Captain William Johnson but DENIES the government's motion to exclude the results of the polygraph examination administered by Dr. Charles Honts." (Smith, 1996, p. 74). After the trial had begun, the Trial Judge granted an appeal from the Government and excluded the polygraph. However, Judge Bowen did not exclude the polygraph for the reasons stated in Abrams (1999). Judge Bowen exercised his discretion under Federal Rule 403 and drew the following conclusion,

> "Nevertheless it is my view that the danger of unfair prejudice and confusion of the issues is clearly presented in the fact of the limited number of all encompassing questions; the restrictions of the opposing party to use substantially the same such questions; and the fact that the questions that were used were limited to, albeit the numerical majority of the case, only one feature of the case. My concern about the length of time, the potential for misleading and confusing the jury and wasting time in expending our trial resources on the polygraph evidence as opposed to other more conventional means to explore the intent and state of mind of the defendant impel me to conclude that this evidence is inadmissible." Bowen (1996, p. 21).

Although he expressed concern about the amount of research on a test that used both directed-lie and probable-lie comparison questions, the judge deferred that issue to the Magistrate and ruled based on a discomfort with the scope of the relevant questions of the examination. Full transcripts from the Gilliard hearing and the complete text of Magistrate Judge Smith's Order and Judge Bowen's ruling are available at http://truth.boisestate.edu.

New Mexico v. Mead. Dr. Raskin tested the defendant. Abrams (1999) stated that the defendant pled guilty after Raskin's cross-examination. The defendant in that case did eventually plead guilty to a lesser charge later in the case, but Abrams (1999) implies that he also admitted his guilt, which is a gross misrepresentation. Mead's lawyer pressured him to enter a guilty plea to a lesser charge because he said he would otherwise be convicted of a much more serious charge. This change in strategy occurred because the prosecution announced that they planned to present a witness from many years ago that would say negative things about Mead that might make the jury believe the charges. However, Mead refused to admit guilt and never confessed, in spite of the judge's orders to do so. The judge then sentenced Mead to an additional four years in prison because he refused to admit that he committed any of the acts to which he pled guilty. That is certainly a far cry from Abram's description that Mead confessed and described the acts in detail.

Utah v. Hofmann. The Hofmann case represents an actual false negative error and was included as such in the Honts and Raskin (1988) field study. Hofmann was charged with two murders in the Salt Lake City area in 1985. Honts examined Hofmann with a test that contained both probable-lie and directed-lie comparison questions. Hofmann passed the test and later confessed to the murders as part of a plea bargain. As part of the plea agreement, Hofmann agreed to tell all about his many crimes and about the polygraph. Drs. Raskin and Honts interviewed Hofmann at the Utah State Prison about the polygraph examination he had beaten. Hofmann claimed to have used hypnosis and biofeedback to beat the test. His knowledge and 15 years of practice of hypnosis and biofeedback were independently confirmed by Dr. Raskin. This subject and the information he provided were so unusual that one of us published a law journal article about the case (Raskin, 1990). We have never denied that an error was made

in this case, and in fact, Honts appeared on the nationally broadcast NPR radio show *All Things Considered* and discussed the error within a short time of the plea agreement.

The presence of a single error provides little information for the scientific evaluation of polygraph accuracy, although it may make for a sensational sound bite. If the polygraph is in fact 95% accurate as Abrams (1999) suggests, that means for every 100 tests there should be five errors. Given a test that is less than perfect, it is inevitable that errors will be made public. However, that should have no impact on the evaluation of scientific research. No one knows more than we do about the negative impact of the Hofmann case being tried in the media. All of this aside, Abrams' (1999) description of the Hofmann case contains serious misrepresentations. There was almost a year between the polygraph examination and the plea agreement, they did not occur at about the same time as Abrams stated. In recent sworn testimony, Abrams also claimed that the Utah Polygraph Association evaluated the polygraph charts from the Hofmann examination, that they all came up with inconclusive results, and that the results of their evaluation were reported at the American Polygraph Association (Trial Transcript, *U. S. v. Walker*, 1999, p139-140). This simply never happened. The Utah Polygraph Association never reviewed the Hofmann polygraph and thus could not have made a report to the American Polygraph Association. In fact, the Hofmann polygraph charts were blindly reviewed by numerous experts, including a two-time president of the American Polygraph Association and many government and law enforcement examiners in the U.S. and Canada. The vast majority found the charts indicative of truthfulness.

Kwan Mak. Mak was accused of being one of three shooters in a robbery and murder of 16 individuals at a gambling club in Seattle's Chinatown district. Mak denied firing any shots and claimed to have already left the building when the shots were fired. Dr. Raskin conducted a polygraph test of Mak, who failed the test. However, his results showed very little reaction to the shooting questions and larger reactions to the presence question. Raskin informed Mak's attorney, who questioned Mak further about his presence. Mak changed his description slightly, and Raskin retested him. He again failed, but the reactions to the shooting questions were again relatively small. Raskin reported Mak as deceptive, but he stated that the relatively small reactions on the most serious issues of shooting and murder indicated a possibility that he may not have fired any shots even though he may have been present. Raskin never reported that Mak was truthful but merely indicated that a full disclosure by Mak of the extent of his involvement might enable Mak to pass a test on the shooting issues. Abram's description of this case is false, and Mak's death penalty conviction was eventually reversed.

Jeffrey MacDonald. Seven years after his conviction, Dr. Raskin tested Dr. MacDonald in 1986 about the murders. He obtained a clearly truthful result. Since then, a massive amount of witness testimony and physical evidence has been obtained that indicates the validity of MacDonald's original description of his family being murdered by members of a drug-crazed cult, the leader of which later confessed. A book and a nationally-aired documentary support MacDonald's description, and recently court-ordered DNA evidence is being examined from the exhumed bodies of his wife and children.

Commonwealth v. Woodward. This was the high profile case of the English au pair that was accused of killing Matthew Eppen. Dr. Raskin tested Miss Woodward with a DLC test and she produced a truthful outcome. After a hearing under the requirements of Massachusetts' law, the judge denied admissibility. Complete transcripts of the polygraph related affidavits and testimony are available at http://truth.boisestate.edu. Abrams (1999) states that Miss Woodward admitted to the police that she shook the baby and that she dropped him (Abrams reports the sex of the deceased incorrectly as female) on his head. This is a grossly inaccurate presentation of the evidence. Miss Woodward stated that she gently shook Eppen in an attempt to awaken him and that his head may have dropped the approximate distance of the thickness of her hand while she was drying him on a towel after his bath. The medical evidence in that case, much of it not made available to the defense until late in the case, was strongly in

the defendant's favor. Moreover, there was additional evidence that the jury was not allowed to see, including a video of the dead child's mother appearing to coach the surviving sibling to lie about Miss Woodward's treatment of the children. The Judge in Woodward essentially nullified the jury's verdict by reducing the sentence to manslaughter and then releasing Miss Woodward for time served. The Judge's decision to set aside the jury's verdict was upheld on appeal. It seems to the present authors that the outcome of the Woodward case paints the polygraph in a very favorable light, and the facts are contrary to Abrams' misleading presentation.

New Mexico v. House. House was tried three times for vehicular homicide. Two of the trials resulted in mistrials because the juries could not reach a decision. House was convicted in the third trial. House admitted drinking a number of beers prior to the head-on collision while he was driving the wrong way on the freeway. However, House stated that he had become disoriented because of an intense episode of familial hemiplegic migraine, from which he had suffered since childhood. Raskin tested House about his claim of experiencing a migraine and found him clearly truthful. An eminent neurologist and headache expert confirmed his migraine claims, but the third jury convicted him because of the blood alcohol level. Raskin's test did not dispute the alcohol problem, and Abrams presentation was again misleading.

New Mexico v. Wilson. Wilson was a female schoolteacher accused of sexually molesting students and charged with molesting a particular girl. Raskin tested her and found her truthful in general, but inconclusive on the question regarding the specific criminal charge. Because of the inconclusive result, Raskin referred her for further evaluation by a psychologist expert in evaluating sexual abuse allegations. His evaluation and interview of her resulted in a confession on the issue that was inconclusive on her polygraph test. Again, Abrams has misrepresented the facts. The police did not obtain the admission as he claimed, but the psychologist selected by Raskin broke the case. Also, she admitted to molesting only the single child and none of the others, as Abrams implied.

Anderson v. Samrock and the Bernalillo Sheriff's Office. In this civil suit, Anderson alleged that he was severely beaten by Deputy Samrock, who had a history of such acts. Consistent with the other evidence, Raskin's polygraph test clearly showed Anderson to be truthful. However, the deputy's tape recording of the incident was unclear and the jury was unable to hear events that Anderson claimed to have occurred. On that basis, they returned a verdict against the plaintiff. This in no way indicates that the polygraph result was incorrect, as Abrams claims.

Griffith v. Melgaard (1995). This was an Idaho Family Court case. Neither Raskin nor Honts tested any of the parties to this civil matter concerning child custody. In the course of an acrimonious divorce and subsequent custody battle Melgaard had been accused of sexually molesting his daughter. Honts did testify to foundational issues in support of two polygraph examinations that were conducted on Melgaard. One of those tests was a directed-lie test conducted by Dr. Dene Simpson; the other was a probable-lie comparison question test conducted by law enforcement. Mr. Melgaard passed both polygraph examinations. Law enforcement reviewed the case and no criminal charges were ever filed against Melgaard. Dr. Abrams (1999) tested the mother, Griffith using a non-standard test that involved having the subject write a statement and then testing the subject on the validity of the statement. Dr. Abrams did not tape his test. Although he claimed to have run three repetitions of the test questions, copies from only two of those recordings were presented to Melgaard's attorney for review. Numerous other experts testified on the medical and psychological findings. Virtually, all of the expert testimony offered by one side of the case was contradicted by expert testimony from the other side. In the end, the judge allowed custody to remain with the mother. This is not an unusual outcome in a custody case, and it certainly dos not indicate that the two polygraph examinations of Melgaard were incorrect.

U. S. v. Freedman. In this federal bank fraud case, one of the co-defendants Skinner was accused of a lesser role. He was tested by George Slattery, who found him truthful using

a standard Backster Zone Comparison format, not a directed-lie. Raskin reviewed the test and testified at the evidentiary hearing. Contrary to Abrams' claim, Skinner never confessed to the allegations but pled to a misdemeanor regarding bank-reporting requirements.

New Mexico v. Raebuck Raebuck was accused of sexual assault and found truthful by examiner Larry Galbreth using a standard military zone comparison test, not a directed-lie. Raskin reviewed the charts and concurred in Galbreth's opinion. Raebuck was convicted. This proves nothing about polygraph accuracy, nor about directed-lie tests.

Idaho v. Kildare. The defendant was tested by a police polygraph examiner and found deceptive using a probable-lie comparison question test, not a directed-lie. Raskin reviewed the charts and found the test to be inconclusive. Kildare confessed and later retracted his confession. He was found guilty. Again, this tells us nothing about polygraph accuracy, especially directed-lie tests.

Wyoming v. Reno. This was a case from 1984 where Honts tested the defendant with a test that included two probable-lie and one directed-lie comparison question. The defendant was accused of a single incident of molestation of a girl friend of his daughter. Supposedly, Reno entered his daughter's bedroom during the friend's sleepover, crawled over his daughter, molested the friend, and left without disturbing his daughter. The daughter testified that nothing happened. The jury voted for conviction, but the judge gave only a brief sentence that was suspended and fined Reno one dollar. There was never any proof to show that the Honts polygraph of Reno was inaccurate.

New Mexico v. Martins. Successful businessman Martins was accused of a serious environmental law violation involving allegations that he illegally removed and dumped what he knew to be asbestos. Martins denied knowing the material was asbestos and even knowing what asbestos was. Raskin's polygraph indicated he was truthful, and the judge admitted the polygraph evidence. On the eve of trial, Martins agreed to plead to a minor count to avoid trial. He never admitted anything that he denied on the polygraph test. Abrams has provided a misleading description.

4. Chimeras and Other Inventions of Fantasy

Abrams would have us believe that he has presented anecdotal evidence that invalidates the published scientific evidence and shows that the directed-lie test produces a high rate of false negative errors. Toward this end, he has presented a series of anecdotes from cases in order to make the claim that they show the occurrence of such false negative errors. However, there are serious problems with his presentation. As described above, he has grossly misrepresented the facts of every case. In addition, many of the cases involved no directed-lie questions.

It should be noted that only one verified error occurred in all of the cases presented by Abrams. Even assuming that tests by Raskin and Honts produce only 5% false negative errors, on an actuarial basis they should have made dozens of such errors. In spite of years of effort, Abrams was able to locate only one true false negative, a celebrated error on a brilliant criminal who used 15 years of hypnosis and biofeedback experience to beat the test. If Raskin and Honts use techniques as inaccurate as Abrams would have the readership of *Polygraph* believe, he should have been able to find the dozens of false negative errors by Raskin and Honts, not just one in more than 15 years.

In contrast, the only three tests by Abrams that we have seen present a picture of error and poor technique. In the Patricia Hearst case, Abrams conducted a probable-lie test that included an inappropriate question requested by the defense attorneys and then scored her as truthful. Independent review by the other two examiners retained by the same defense counsel clearly indicated a strong deceptive result on that question. In another case, Abrams reported an educational researcher deceptive after testing him with a probable-lie test on sexual assault allegations by a young teacher. The defense attorney who had retained Abrams to conduct the test asked Raskin to review it. Raskin scored the charts as definitely truthful and one week later the

young woman confessed that she had fabricated the entire incident, proving Abrams incorrect. Using the same type of reasoning taken by Abrams with his anecdotal evidence about the directed-lie, based on Abrams use of the probable-lie, that type of test produces 100% errors.

Abrams (1999) uses a term, hybrid directed-lie test, which has no operational definition other than as a general reference to the field practices of Dr. Honts and Raskin. Over the years he has used this term inconsistently. When originally used in the Gilliard case it referred to a test that used both probable-lie and directed-lie comparison questions. In the 1999 paper Abrams appears to have expanded this contrivance to include tests run by Dr. Raskin and Honts that use only directed-lie tests. The motives for this change of definition are not clear, but it is sure to cause confusion in those who read the Gilliard (1998) appellate decision that concerns a test that includes both types of comparison questions.

Abrams (1999) then attempts to establish by edict that the test currently used by Drs. Raskin and Honts in their field practices is different from the test used by the U. S. Government. However, those differences are never explicated, except for a statement that Drs. Honts and Raskin have discussions with their subjects between question repetitions and U. S. Government examiners do not. Given the quote from Fuse (1982) reproduced above, it is clear that even that trivial difference is inaccurate for at least some of the directed-lie tests run by the Government. In the present paper and elsewhere (Honts, 1999), we have clearly shown that the preponderance of the scientific data indicate that between-chart discussions increase test accuracy even when they favor the comparison questions.

The American Heritage Dictionary defines chimera as follows: "An imaginary monster made up of grotesquely disparate parts." (1992, p. 332) In his abstract Abrams (1999) claims to have " . . . provided evidence supporting his view that this technique [the DLC] neither should be admitted into court nor employed as a polygraph technique . . ." (p. 223). He did nothing of the sort. Abrams (1999) provides no new scientific data. Rather, he misrepresents much of the existing research, and he argues by anecdotes that are frequently inaccurate. He has, in effect, created a chimera. If the polygraph profession is to progress, it must rely on scientifically obtained research data and results. Readers interested in the science concerning the DLC and the review of questions between repetitions are referred to Honts (1999), Honts and Gordon (1999) and to Raskin et al., (1997). Nothing Abrams has written has any impact on the scientific validity of the research reviewed in those reports.

References

Abrams, S. (1991). The directed-lie control question. *Polygraph*, 20, 26-31.

Abrams, S. (1999). A response to Honts on the issue of the discussion of questions between charts. *Polygraph*, 28, 223-229.

Beardsley, T. (1999). Truth or consequences. *Scientific American*, 281(4), 21-24.

Bradley, A. R., & Wood, J. M. (1996). How do children tell? The disclosure process in child sexual abuse. *Child Abuse & Neglect*, 20, 881--891.

Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993). 509, U. S. 579, 113 S. Ct. 2786.

Dawson, M. E. (1981). Physiological detection of deception: Measurement of responses to questions and answers during countermeasure maneuvers. *Psychophysiology*, 17, 8-17.

Department of Defense Polygraph Institute Research Division Staff (1997). A comparison of psychophysiological detection of deception accuracy rates obtained using the Counterintelligence Scope Polygraph (CSP) and the Test for Espionage and Sabotage (TES) question formats. *Polygraph*, 26, 79-106.

Department of Defense Polygraph Institute Research Division Staff (1998). Psychophysiological detection of deception accuracy rates obtained using the Test for Espionage and Sabotage (TES). *Polygraph*, 27, 68-73.

Dollins, A.B. (1998). A guide to Department of Defense Polygraph Institute research interests. *Polygraph*, 27, 89-95.

Fuse, L. S. (1982). Directed-lie control testing technique. Unpublished manuscript.

Griffith v. Melgaard. (1995) Idaho District Court for Ada County.

Honts, C. R. (1999). The discussion of questions between list repetitions (charts) is associated with increased test accuracy. *Polygraph*, 28, 117-123.

Honts, C. R., & Gordon, A., (1998). A critical analysis of Matte's analysis of the directed lie. *Polygraph*, 27, 241-252.

Honts, C. R., & Perry, M. V. (1992). Polygraph admissibility: Changes and challenges. *Law and Human Behavior*, 16, 357-379.

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., 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.

Lamb, M. (1994). The investigation of child sexual abuse: An interdisciplinary consensus statement. Published simultaneously in, *Child Abuse & Neglect*, 18, 1021-1028; Expert Evidence, 2, 151-156; *Family Law Quarterly*, 28, 151-162; *Journal of Child Sexual Abuse*, 3, 93-106; and *Scandinavian Journal of Social Welfare*, 3, 175-180.

Office of Technology Assessment (1983). Scientific validity of polygraph testing: A research review and evaluation. OTA-TM-H-15. Washington, DC: U. S. Congress, Office of Technology Assessment.

Patrick, C. J. & Iacono W. G. (1989). Psychopathy, threat, and polygraph test accuracy. *Journal of Applied Psychology*, 74, 347-355.

Raskin, D. C. (1990). Hofmann, hypnosis and the polygraph. *Utah Bar Journal*, 3(9), 7-10.

Raskin, D. C., Honts, C. R., & Kircher, J. C. (1997). The scientific status of research on polygraph techniques: The case for polygraph tests. Chapter in, D. L. Faigman, D. Kaye, M. J. Saks, & J. Sanders (Eds.) *Modern scientific evidence: The law and science of expert testimony* (pp. 565-582).

Reed, S., (1994). A new psychophysiological detection of deception examination for security screening. *Psychophysiology*, 31, S80. (Abstract)

Summit, R. C. (1983). The child sexual abuse accommodation syndrome," *Child Abuse and Neglect*, 7, 177-193.

Summit, R. C. (1992). Abuse of the child sexual abuse accommodation syndrome. *Journal of Child Sexual Abuse*, 16.

U. S. v. Galbreth. (1995). 908 F. Supp. 877.

U. S. v. Gilliard (1996). Transcript of the Daubert hearing. Available online: http://truth.idbsu.edu/polygraph/gilliard/abrams.htm

U. S. v. Gilliard (1998). United States Court of Appeals, Eleventh Circuit. No. 96-9459. Available online: http://www.law.emory.edu/11circuit/wpds/jan98/96-9459.man

U. S. v. Walker (1999). United States District Court for Alaska, Case no. A98-0158-CR (JKS). Hearing transcript available at, http://truth.boisestate.edu.

The Frequency of Appearance of Evaluative Criteria in Field Polygraph Charts

Norman Ansley and Donald J. Krapohl

Abstract

Every appearance of each of 22 response patterns considered to be diagnostic for the detection of deception by the US Department of Defense Polygraph Institute (DoDPI) was tabulated for 177 cases (616 polygraph charts) selected from the DoDPI database of confirmed field cases. The sets of charts were in 16 different formats, but all were a form of zone comparison. We found the total number of appearances of these criteria ranged from 5 to 4,793. A rank ordering by frequency of the 22 criteria stayed remarkably constant across questions, gender, and truthful or deceptive status. There was a reduction in the number of reactions in the second and third charts of nondeceptive examinees in all three physiological channels and a similar reduction in the electrodermal and cardiograph channels of deceptive examinees. However, the respiratory pattern showed an increase in reactions in successive charts of deceptive examinees. We also found more reactions and a higher tonic heart rate for the deceptive than the nondeceptive examinees. The 1,780 relevant question presentations produced 6,453 reactions, for an average of 3.6 reactions per question. The 1,932 comparison questions produced 6,777 reactions, for an average of 3.5 per question. The technical questions (irrelevant, symptomatic, sacrifice relevant) were asked 2,154 times and produced 7,484 reactions for an average of 3.5 per question. The pneumograph produced 19% of the reactions, the cardiograph 26%, and the electrodermal 55%.

Key words: cardiovascular, deception criteria, distribution of reactions, electrodermal responses, habituation, normative data, polygraph tracing features, pneumograph, tonic heart rate, zone comparison formats.

Over the last 75 years, lists of diagnostic polygraph tracing features have evolved from the observations of examiners in the conduct of countless field examinations. Polygraph schools over the years incorporated those observations into their curricula, developed scoring rules for them, and the instruction influenced the chart interpretations of generations of new polygraph students. While most of the instruction regarding the diagnostic features in polygraph tracings are shared among different schools, surprisingly little work has been done regarding the frequency and predictive value of those reaction criteria. We know from Capps & Ansley (1992) the types of polygraph tracing features examiners use in their analyses of the charts, but that study did not tell us what reactions were present in the tracings but not used. It would be of interest to explore the incidence of polygraph features in field cases, separate from their diagnostic use.

The US Department of Defense Polygraph Institute teaches that there are 23 specific features in polygraph tracings that are used in numerical analysis. Twelve of the features are found in the two pneumograph channels, three in the electrodermal channel, and eight in the cardiograph channel. These criteria were previously reported by Swinford (1999), and are reprinted here.

Acknowledgements

This article is based on a larger study funded by the US Department of Defense Polygraph Institute. The full study can be requested through the Defense Technical Information Center, 8726 John J. Kingman Road, STE 0944, Fort Belvoir, VA 22060-6218. The study was prepared under ONR Grant Number N00014-98-1-0863.

The senior author is Editor Emeritus of the American Polygraph Association, President of Forensic Research, Inc., and a former federal examiner. The junior author is a federal polygraph examiner and researcher, and current editor of the journal *Polygraph*. Requests for reprints can be sent to: Norman Ansley, 35 Cedar Road, Severna Park, MD 21146-3715.

Respiration
R1. Rate decrease
R2. Rate increase
R3. Inhalation/exhalation ratio change
R4. Amplitude increase
R5. Amplitude decrease (suppression)
R6. Progressive increase or decrease
R7. Progressive increase and return
R8. Progressive decrease and return
R9. Baseline change – temporary
R10. Baseline change – permanent
R11. Apnea – holding (inhalation)
R12. Apnea – blocking (exhalation)

Electrodermal
E1. Amplitude change
E2. Complex response
E3. Response duration and return

Cardiovascular
C1. Baseline increase and decrease
C2. Baseline increase
C3. Baseline decrease
C4. Amplitude increase
C5. Amplitude decrease
C6. Rate increase
C7. Rate decrease
C8. Premature ventricle contractions

The purpose of the present paper is to look at the incidence of the DoDPI reaction criteria in field cases. In addition to generic normative data regarding the frequency of reactions, we wanted to know if the distribution of criteria differed between deceptive and nondeceptive cases. We were also interested in evidence of habituation of responding across successive charts, or within charts across questions. Finally, we wanted to know if the tonic heart rate of deceptive examinees was different from the heart rate of truthful examinees. Though most of the DoDPI reaction criteria have been used in field practice and various schools of instruction for about 50 years, reports of normative field data are sparse. A modest investigation of cardiograph responses was reported by Jensen (1981), and his results were compared to the present findings.

Method

Cases

A total of 177 polygraph cases were selected at random from the DoDPI database of confirmed cases by the junior author. All cases had been conducted in the field by federal or law enforcement polygraph examiners using Axciton computer polygraphs (Axciton Systems, Houston, TX). The only criteria for selection of cases were that they be identified as single-issue field zone comparison examinations. For the 161 cases where gender was identified, there were 115 men and 46 women. There were 111 deceptive and 66 nondeceptive cases. Of the 115 males, 71 were deceptive, and 44 nondeceptive. Among the 46 women, 31 were deceptive and 15 were nondeceptive. For the remaining 16 cases where gender was not recorded, 9 were deceptive and 7 were nondeceptive.

Human evaluator

The frequency counts were performed by the senior author, who has 49 years of polygraph experience. He was blind to ground truth and gender until the tabulations were complete.

Tabulation procedure

The list on Forensic Research, Inc. (FRI) Form 1 (Appendix A), and the definitions of scoring criteria used in this study are from the DoD Polygraph Institute. However, FRI Form 1 deleted the premature ventricle contraction (PVC) criterion because it is not generally deemed an autonomic response. Moreover, of the 5,866 question presentations in these cases, PVCs occurred 30 times, of which 18 were in one set of charts. Given the low incidence, they were not considered further.

An FRI Form 1 was made for each chart. The experienced examiner noted the presence of each of the criteria for each question presented on each chart. The data were then tabulated, and sorted for type of question, ground truth (deceptive or nondeceptive), gender, and polygraph channel.

Results

The 1,780 relevant questions produced 6,453 reactions, for an average of 3.6 per question. The 1,932 comparison questions produced 6,777 reactions, for an average of 3.5 per question. The 2,154 technical questions (irrelevant, sacrifice relevant, symptomatic) produced 7,484 reactions, for an average of 3.5 per question. In terms of types

of questions, 31% were relevant, 33% were comparison, and 36% were technical. Of the 20,714 reactions, 3,848 or 19% were from the pneumograph, 11,414 or 55% were from electrodermal, and 5,452 or 26% were from the cardiograph. It is interesting to note that the percentages shown here are similar to the weights give by some scoring algorithms.

The number of times each of the 22 reaction types appeared is on Table 1. Next to the number is the percentage of the total reactions the number represents. At the top is electrodermal amplitude change (E1) which appeared 4,793 times. E1 accounted for 26% of the total appearance of all reactions from all polygraph channels. At the bottom is cardiograph amplitude increase (C4) which appeared only five times, or less than one-half of one percent.

Table 1. Reaction criteria ranked by frequency.

Criterion	Description	Frequency	Percent
E1	Amplitude change	4793	26
E3	Duration	4496	24
C1	Baseline increase & decrease	2778	15
E2	Complex response	1051	6
C5	Amplitude decrease	940	5
R4	Amplitude increase	704	4
R9	Baseline change - temporary	683	4
C2	Baseline increase	578	3
R5	Amplitude decrease/suppression	476	3
C3	Baseline increase	400	2
RI0	Baseline change - Permanent	389	2
R1	Rate decrease	318	2
R8	Progressive decrease & return	265	1
R12	Apnea - (exhalation)	182	1
R2	Rate increase	154	1
R7	Progressive increase & return	107	1
R6	Progressive increase/decrease	102	1
R3	I/E Ratio change	62	less than .5%
C6	Rate increase	25	less than .5%
C7	Rate decrease	23	less than .5%
R11	Apnea - Holding (inspiration)	9	less than .5%
C4	Amplitude increase	5	less than .5%

Codes
P=pneumograph
E=electrodermal
C=cardiograph

Number = criterion number. See FRI Form 1 in Appendix A

Table 2. Ranking of response criteria by frequency and deceptiveness status.

	All Cases		Nondeceptive Cases		Deceptive Cases	
Rank	Criterion	Percent	Criterion	Percent	Criterion	Percent
1	E1	26	E1	25	E1	26
2	E3	24	E3	24	E3	24
3	C1	15	C1	15	C1	15
4	E2	6	E2	6	C5	5
5	C5	5	R5	4	E2	4
6	R4	4	R4	3	R9	4
7	R9	4	R9	3	R4	4
8	C2	3	C2	3	C2	3
9	R5	3	C3	3	R53	3
10	C3	2	R5	3	C3	2
11	R10	2	R10	3	R10	2
12	R1	2	R8	2	R1	2
13	R8	1	R1	1	R8	1
14	R12	1	R7	1	R12	1
15	R2	1	R2	1	R2	1
16	R7	1	R6	1	R6	1
17	R6	1	R12	0	R3	0
18	R3	0	R3	0	R7	0
19	C6	0	C7	0	C7	0
20	C7	0	Rll	0	R4	0
21	R11	0	C4	0	C4	0
22	C4	0	C6	0	C6	0

In 1981 Carl W. Jensen published a study entitled "Frequency of occurrence of specific reaction criteria as observed in the cardio tracing." When the terminology of Jensen's study is matched with DoDPI's, and both data sets are ranked by frequency, the lists are strikingly similar (Table 3). This finding is reassuring for two reasons. First, the data from each study support the other, lending credibility to both. Second, Jensen's data were produced by analog instruments, and the present data were recorded digitally. The highly similar outcomes of the two studies suggest that the output signals from the two recording instruments have much in common, and may alleviate concerns in some quarters that the cardiograph tracings of computer polygraphs are different in a significant way from the older analog instruments.

Table 3.
Frequencies of cardiograph criteria for the present data, and the Jensen study (1981).

Present Study		Jensen's Study	
Criteria	Frequency	Criteria	Frequency
Baseline increase & decrease	2778	Baseline increase & decrease	363
Pulse amplitude decrease	940	Pulse amplitude decrease	326
Baseline increase	578	Baseline increase	172
Baseline decrease	400	Pulse amplitude increase	52
Pulse rate increase	25	Baseline decrease	48
Pulse rate decrease	23	Pulse rate increase	43
Pulse amplitude increase	5	Pulse rate decrease	20

We looked at the serial effects of questions by deceptiveness and non-deceptiveness. See Tables 4, 5, and 6 for frequencies and proportions of reactions in the cardiograph, electrodermal, and respiratory channels, and the summary on Table 7. There was a consistent reduction of reactions in the second and third charts of nondeceptive examinees in all three channels compared to the first chart. However, the respiratory pattern showed an increase in reactions in successive charts of deceptive examinees. The unusual effect can be seen in eight of the ten questions on Table 6. The underlying cause of this anomaly warrants further study.

Table 4. Change in the frequency of DoDPI diagnostic features in the cardiograph by question and by chart.

	Deceptive Cases (n=111)					Nondeceptive Cases (n=66)				
	Chart 1	Chart 2	change	Chart 3	total change	Chart 1	Chart 2	change	Chart 3	total change
Question										
1	105	106	1	77	-28	61	71	10	55	-6
2	114	112	-2	99	-15	67	64	-3	56	-11
3	103	101	-2	98	-5	83	72	-11	50	-33
4	98	119	21	122	24	77	63	-14	50	-27
5	136	129	-7	101	-35	68	62	-6	63	-5
6	96	103	7	105	9	72	65	-7	60	-12
7	134	113	-21	123	-11	74	59	-15	40	-34
8	79	87	8	83	4	61	50	-11	44	-17
9	105	93	-12	76	-29	59	133	74	51	-8
10	106	100	-6	100	-6	59	28	-31	26	-33
Total	1076	1063	-13	984	-92	681	667	-14	495	-186

Table 5. Change in the frequency of DoDPI diagnostic features in the electrodermal channel by question and by chart.

	Deceptive Cases (n=111)					Nondeceptive Cases (n=66)				
	Chart 1	Chart 2	change	Chart 3	total change	Chart 1	Chart 2	change	Chart 3	total change
Question										
1	222	210	-12	191	-31	136	136	0	120	-16
2	234	213	-21	216	-18	155	137	-18	120	-35
3	231	221	-10	190	-41	172	131	-41	94	-78
4	185	199	14	193	8	141	129	-12	121	-20
5	223	229	6	224	1	155	125	-30	141	-14
6	247	215	-32	209	-38	180	136	-44	132	-48
7	234	230	-4	214	-20	163	120	-43	111	-52
8	203	190	-13	192	-11	124	110	-14	97	-27
9	225	189	-36	149	-76	136	175	39	114	-22
10	217	185	-32	182	-35	128	60	-68	48	-80
Total	2221	2081	-140	1960	-261	1490	1259	-231	1098	-392

Table 6. Change in the frequency of DoDPI diagnostic features in the pneumograph channel by question and by chart.

	Deceptive Cases (n=111)					Nondeceptive Cases (n=66)				
	Chart 1	Chart 2	change	Chart 3	total change	Chart 1	Chart 2	change	Chart 3	total change
Question										
1	56	54	-2	60	4	38	41	3	38	0
2	63	73	10	77	14	46	42	-4	47	1
3	77	86	9	89	12	48	33	-15	40	-8
4	59	86	17	92	13	56	43	-13	38	-18
5	82	84	2	97	15	47	38	-9	37	-10
6	81	88	7	91	10	50	44	-6	34	-16
7	75	87	12	93	18	47	32	-15	38	-9
8	76	78	2	83	7	41	29	-12	33	-8
9	73	70	-3	59	-14	34	7	-27	37	3
10	83	55	-28	61	-22	41	26	-15	22	-19
Total	725	761	26	802	57	448	335	-113	364	-84

Table 7. Reaction totals by channel and chart.

	Deceptive Cases (n=111)			Nondeceptive Cases (n=66)		
	Chart 1	Chart 2	Chart 3	Chart 1	Chart 2	Chart 3
Respiration	745	761	802	448	335	364
Electrodermal	2221	2081	1960	1490	1259	1098
Cardiovascular	1076	1063	984	681	667	495
Average	1347.3	1301.7	1248.7	873.0	753.7	652.3

Tonic Heart Rates

From Table 8, we see average tonic heart rates of deceptive examinees were faster than the tonic rates of nondeceptive examinees, and the difference was significant ($z=2.87$, $p<.05$). The pattern held true for men and women, and at the beginning and the end of charts.

Table 8. Heart beats per minute for men and women during polygraph testing.

	Deceptive		Nondeceptive	
	Men	Women	Men	Women
Beginning	89	98	84	91
End	88	97	84	91

Discussion

The principal purpose of this study was to develop normative data for the DoDPI evaluative criteria in field polygraph charts. Two findings of the study are worthy of special note and comment. First, when the evaluative criteria are placed in rank order by frequency of appearance, it is apparent that some appear very rarely. Given the low incidence of some criteria, the present writers suggest that the list of evaluative criteria could be shortened to some extent without hampering day-to-day chart interpretation. For example, amplitude increase (C4) could be combined with amplitude decrease (C5) as simply cardiograph amplitude change. The criteria of pulse rate increase (C6) and rate decrease (C7) could be dropped entirely, as they each constituted less than one-half of one percent of all responses. However, if instrument manufacturers would add a cardiotachometer as an optional feature, these criteria might have utility. Some of the automated algorithms do make good use of this measure. The present findings with respect to pulse rate may simply reflect the difficulty in discerning subtle rate changes with the instrumentation used here. With regard to the respiration channel, experience with older instruments suggests that the inhalation/exhalation ratio (R3) might be more prevalent than was evident with these digitized instruments. If computer instruments do not manifest more inhalation-exhalation ratio changes than appeared here, the criterion might be considered for deletion. Apnea - holding (R11) could be combined with apnea - blocking (R12), as just apnea. One more combination of respiration criteria would make sense; merge baseline change - temporary (R9) with baseline change – permanent (R10). The differences between the two are not always clear, and the distinction does not appear to add to the probative value of the test.

The second noteworthy finding was the shrinking number of reactions across successive charts, suggesting the influence of generalized habituation. This was not an unexpected finding. However, we did not anticipate the increasing number of respiration reactions over charts that occurred exclusively with deceptive cases. The reason that respiration responses for deceptive examinees ran counter to the trend of habituation found for all other channels for both deceptive and nondeceptive examinees is beyond the scope of this study. Moreover, such a pattern would not be predicted from the published literature on polygraphy. If the finding is confirmed in other research, it may point to an unresolved area in polygraph theory.

The normative data in this paper are a small part of a study conducted by the first author for DoDPI. Those interested in the complete report should contact the Defense Technical Information Center, 8726 John J. Kingman Road, STE 0944, Fort Belvoir, VA 22060-6218. The study was prepared under ONR Grant Number N00014-98-1-0863.

References

Capps, M.H., & Ansley, N. (1992). Numerical scoring of polygraph charts: What examiners really do. *Polygraph*, 21(4), 264-320

Jensen, C.W. (1981). Frequency of occurrence of specific reaction criteria as observed in the cardio tracing. *Academy Journal*, 4(2), 5-7.

Swinford, J. (1999). Manually scoring polygraph charts utilizing the seven-position numerical analysis scale at the Department of Defense Polygraph Institute. *Polygraph* 28(1), 10–27.

Data Evaluation Form 1

Forensic Research, Inc. 1998

File Name:_____ Chart:____of____ Test Format:_____

Heart Rate - Beginning:_____End:_____ PVCs useful? Yes No Other:_____

	1	2	3	4	5	6	7	8	9	10	T	TC	TR
RESPIRATION													
1. Rate Decrease													
2. Rate Increase													
3. I/E Ratio Change													
4. Amplitude Increase													
5. Amplid Decrease/Suppression													
6. Progressive Increase/Decrease													
7. Progressive Increase & Return													
8. Progressive Decrease & Return													
9. Baseline Change - Temporary													
10. Baseline Change - Permanent													
11. Apnea - Holding (inspiration)													
12. Apnea - Blocking (Exhalation)													
ELECTRODERMAL													
1. Amplitude Change													
2. Complex Response													
3. Response Duration & Return													
CARDIOVASCULAR													
1. Baseline Increase & Decrease													
2. Baseline Increase													
3. Baseline Decrease													
4. Amplitude Increase													
5. Amplitude Decrease													
6. Rate Increase													
7. Rate Decrease													
8. P.V.C.													

Gender: Male Female Race: Caucasian African American Hispanic Asian Age:_____

Comments:_____

Appendix A

Guide for Performing the Objective Scoring System

Donnie W. Dutton

Key words: Objective Scoring System

Imagine you are sitting in your office, when the phone rings. It's the Assistant U.S. Attorneys Office, and she wants you to testify on Friday about that polygraph test you did some months ago. You pull the file, and upon reviewing it you realize that you never got that second opinion that you were going to get. Not only is that second opinion important, but having an empirical foundation for the analyses that support your conclusion would go a long way toward aiding your credibility in court.

The Objective Scoring System (OSS) is one option for the polygraph expert in this circumstance. The OSS is a manual numerical scoring method developed specifically for evidentiary applications. It allows virtually perfect agreement among multiple scorers, and it is possible to estimate decision accuracy from those scores. While the OSS uses far fewer criteria for evaluating data than other systems, it requires scrupulous attention to the scoring protocol to deliver these advantages over traditional 7-position scoring. Despite the value of the OSS for evidentiary purposes, few field practitioners are proficient with this scoring technique. Previous reports of the OSS (Krapohl & McManus, 1999; Krapohl & Norris, 2000) demonstrated the relative accuracy of the OSS to traditional 7-position scoring, but neither provided step-by-step instructions of the scoring method that would permit practitioners to use the OSS in the field. To date only those examiners receiving special training are sufficiently knowledgeable of the method to employ it. This paper is intended as an instruction guide for performing the OSS.

Before taking up the OSS protocol, it is important to state three disclaimers. First, OSS was based on field examinations consisting of three-charts, each chart containing three relevant and comparison questions, conducted as a single-issue DoDPI Zone Comparison Technique (DoDPI, 1992), using exclusionary probable-lie comparison questions. While the OSS may tolerate some deviations from those conditions, its validity is not established for other formats or configurations at this time. Second, the scoring method assumes that the data are adequate: highly unstable, heavily artifacted, or very unresponsive tracings that are not suitable for manual scoring are also not suitable for the OSS. Similarly, examinee manipulations of the tracings may preclude the use of the OSS. Like any scoring methodology, the OSS cannot compensate for inadequate data, poor instrumentation, or bad technique. Third, the OSS takes much longer to perform than traditional 7-position scoring, with an average of 45 minutes per case. Consequently, it is not for everyday use. It is designed exclusively for evidentiary applications, or other settings where interrater agreement must be virtually perfect, and where error probability must be established. Traditional scoring works very well in the field, but it does not offer these two capabilities. The OSS fills a special niche among scoring methodologies, but it is not appropriate for daily use.

The Tools

If the charts meet the requirements for use of the OSS, users will need tools for performing the measurements of the tracing

The author is a federal examiner and APA Director. The statements made in this paper do not necessarily represent the views of the Department of Defense or the U.S. Government. Reprint requests should be sent to Donnie Dutton, P.O. 10411, Ft. Jackson, SC 29207.

features. Since the features in the electrodermal and blood volume tracings are straightforward measurements of amplitude, any ruled apparatus will work. Some polygraph manufacturers sell transparent plastic overlays that permit examiners to measure features in millimeters and time in seconds, these are well suited for working with the OSS.

The sigmoid waveforms of the respiration calls for a special device that can measure features in curved tracings. Obviously, a straight ruler will not suffice. To accurately measure the respiration tracing features, some OSS users have purchased a commonly available device called a planimeter. The handheld planimeter is often used to measure road distances on maps, and one can be purchased at almost any department or automotive store. When choosing a planimeter, select one that has the best resolution, that is, the one that can record the smallest units of distance. This is because precision is an important quality of the OSS, and there could be a decline in effectiveness when measurements are imprecise or unreliable. A very desirable feature to watch for in a planimeter is a counter that continues to add the distances forward, even when the roller is moved backwards. This feature helps the user avoid awkward hand positions while tracing the respiration waveform.

The Features

The OSS uses only one feature each for the polygraphic channels of electrodermal, blood volume and respiration. Collectively, these are called the "Kircher features", from the work of Kircher and Raskin (1988) to identify the diagnostic features in polygraph tracings. For the electrodermal channel, it is the peak amplitude for a phasic response that begins from 0.5 seconds to about eight seconds after question onset. If an electrodermal response (EDR) begins before 0.5 seconds after question onset, it cannot be attributed to the test question. Similarly, examiners should be suspicious of responses that begin after eight seconds from question onset, unless there are defensible reasons to accept them as genuine.

The feature in the blood volume tracing is the increase taking place from question onset to the end of the response, or the end of the question window. If the tracing drops downward at stimulus onset and never rises above the level at question onset, it is recorded as having 0 units of amplitude, since negative amplitudes are not used in OSS.

Of the three Kircher features, the respiration channel produces the one least intuitive to polygraph practitioners. It is called respiration line length (RLL)(Timm, 1982). RLL in the Kircher features ensemble is the length of 10 seconds of respiration tracing if it were straightened into a line. In order to obtain this measurement, OSS users must use the planimeter or similar device to determine how long the respiration tracing is in the 10-second window beginning at question onset. RLL captures both respiratory suppression, and increases in the inhalation/exhalation (I/E) ratio, in a single value. Each of these responses causes a shortening of the RLL. Short RLLs are associated with physiological arousal, and are indicative of deception in conventional polygraphy. Figure 1 shows the Kircher features.

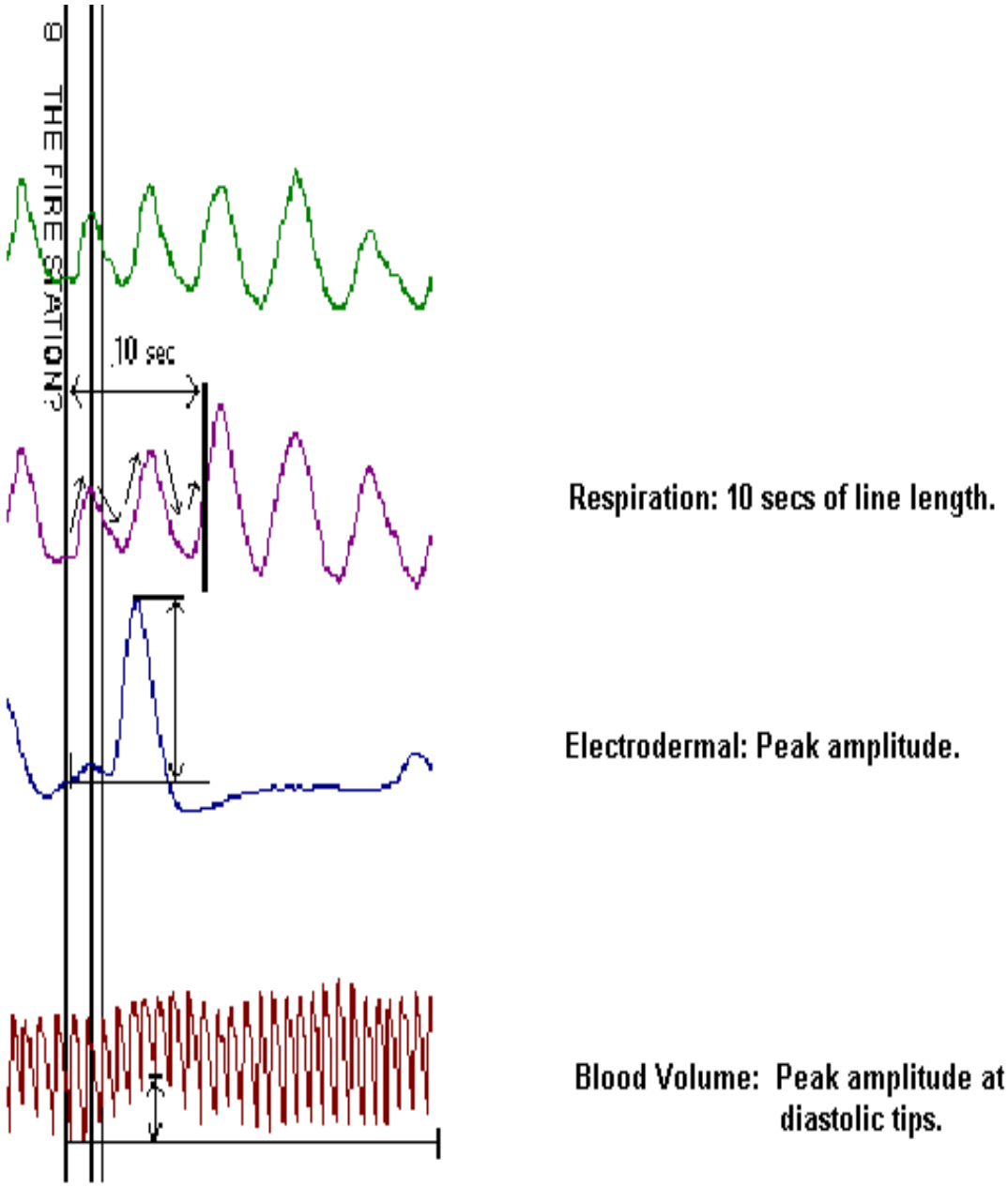
The Measurements

Respiration Line Length

First, it is important to mark the 10-second window for measurement. The window begins at question onset, a point that is automatically marked on computer polygraphs, but may require manual demarcation on charts produced with analog instruments. From the question onset, mark the respiration waveform at 10 seconds. It is important that the waveform be measured precisely. With the planimeter, or similar device, trace the respiration waveform for the entire 10-second window of each relevant and comparison question. If there are two respiration channels, it will be necessary to measure both unless one does not meet technical standards. Place the measurement in the proper place on the OSS data sheet found in Appendix A. For example, if the upper respiration of the first comparison question measured 22.6 units, this number would be written in the first cell in the upper left corner, where the column marked CQ1 and row RLL 1 meet. If the measurement had been for the lower respiration channel, the cell just below this one would be used, etc. Each RLL for each

relevant and comparison question must be measured, and that measurement recorded in the OSS data sheet under the heading Measurements. Artifacted responses are marked with an A, to indicate that the tracing was not scorable.

Figure 1. How the Kircher features are measured.



Electrodermal Response

In fortunate contrast to the measurements of the respiration channel, the EDR measurements are quite easy. The diagnostic feature in electrodermal activity [EDA] channel is the amplitude of the phasic response. It is measured from question onset to the peak of the response. The examiner must be confident that the phasic electrodermal response is associated with the test question. Therefore, responses that begin before the question presentation should be ignored. Also, it becomes increasingly unlikely that a phasic response was evoked by the test question the later it begins after about eight seconds. Due to some examinees characteristically responding late, a firm timeframe for phasic response onset cannot be stated absolutely.

The amplitude measurements for the EDRs should be placed in the EDR column of the data sheet. There is one EDR measurement for each question. As with the respiration channel, artifacts are denoted with an A in the data sheet, to indicate that the tracing was not scorable.

Blood Volume (BV)

Measuring the amplitude of the blood volume is quite similar to measuring the EDA phasic response. Examiners can draw a line that averages the systolic and diastolic points, or more conveniently, this rise can be tracked at the diastolic tips. The initial measurement point is at question onset, and the window extends to the greatest amplitude occurring before the presentation of the next question. It may be easier to draw a horizontal line forward from the reference point at question onset to help make the amplitude measurement. As noted earlier in the Electrodermal Response section, the further away from question onset, the more unlikely that response is to be associated with phasic response. Spontaneous responses, that is, those not associated with the test question, would not be included in the measurement. The amplitude measurement would be entered in the BV row for each question under the heading Measurements. As is the case in the other channels, artifacts are denoted with an A in the data sheet, to indicate that the tracing was not scorable.

The Computations

All OSS scores are based on the ratio of the Kircher features for the relevant question divided by those of the comparison question (R/C). For example, if the BV amplitude were 34 units on the relevant question, and the BV amplitude for the comparison question was 23 units, you would divide 34 by 23, yielding a ratio of 1.48. Round all ratios to two decimal points. If the RLL for a relevant question were 28 units, and the RLL for the comparison question were 46 units, the ratio would be 0.61 (28/46). Similarly, if the amplitude of the EDR to the relevant question were 16 units, and the EDR to the comparison question were 33 units, the ratio would be 0.48 (16/33). Ratios must be calculated for every relevant/comparison question pairing, and entered in the data sheet under Ratios (R/C). For each of the three spots, there would be a ratio for the two RLLs, the EDRs, and the BVs. The significance of these ratios will be explained in the next section.

There are two significant hurdles to overcome with the computations: artifacted tracings, and measurements of 0. Artifacts usually interfere with the ability to assign a score to a channel. There are some exceptions. If the response occurs on any single comparison question, it is permissible to use the next closest comparison question on the chart for evaluation. The same is not true with an artifact on relevant question, however. When this occurs, the score must be a 0. With regard to amplitude measurements of 0 in the EDA and BV, they become a problem if they occur on any comparison question. This is because ratios are calculated as R/C, and a zero in the denominator of a fraction has no mathematical meaning. If an EDR or BV response has no amplitude, or a negative amplitude, it is permissible to enter a 0.01 value in the cell for the measurement.

Scoring rules of different formats dictate which relevant questions are used with which comparison questions. The data used for the development of the OSS was the DoDPI ZCT. According to the scoring rules of the DoDPI ZCT, question R7 is compared to C6, R10 to C9, and question R5 is compared to the stronger of either C4 or C6. The stronger of

C4 and C6 would be greater BV and EDA amplitude, and shorter RLL.

The Scores

Once the ratios have been computed and entered on the data sheet, scores can be assigned. The individual scores are based on the ratios of R/C. At the bottom of the OSS data sheet is a table, reprinted here as Table 1. Using the examples from above, the ratio of the BV response for that question was 1.48. Referring to Table 1, a ratio of 1.48 in the BV would be assigned a score of -2, since the ratio falls between 1.66 and 1.30. For the RLL, the ratio was 0.61, which corresponds with a score of -3. Note that, though there are two respiration channels, only one score is used at each spot for respiration; the one score that is farther from 0. If the scores for the two RLL ratios are of opposite sign, (+ and -) a score of 0 is assigned.

For the EDR, the ratio in the example was 0.48. The score assigned to an EDR ratio of 0.48 is a +4. Note that the EDA channel is weighted: Instead of +/- 1, 2 and 3, the scores are +/- 2, 4 and 6. Weighting the EDA in this fashion increases the accuracy of the OSS.

Table 1. Table for conversion of ratios to scores in the Objective Scoring System.

Channel	Scoring Table						
RLL	0.00 - 0.79	0.80 - 0.89	0.90 - 0.96	0.97 - 1.03	1.04 - 1.10	1.11 - 1.25	1.26 - 999
score =>	-3	-2	-1	0	1	2	3
EDR	999 - 2.45	2.44 - 1.61	1.60 - 1.21	1.20 - 0.93	0.92 - 0.68	0.67 - 0.44	0.43 - 0.00
score =>	-6	-4	-2	0	2	4	6
BV	999 - 1.67	1.66 - 1.30	1.29 - 1.06	1.05 - 0.89	0.88 - 0.72	0.71 - 0.54	0.53 - 0.00
score =>	-3	-2	-1	0	1	2	3

This process of assigning scores to ratios is continued, until every channel in every spot for all three charts has received a score. These scores are tallied for each spot, and the spots are summed for a grand total. This total is used to render a decision from the physiological data.

An inconclusive region of +/-7 is recommended with the OSS, as the empirical evidence suggests that the proportion of errors will be about 0.05 for both deceptive and nondeceptive examinees. All totals of -8 or lower would be called Deception Indicated (DI), and those +8 or greater would be No Deception Indicated. However, examiners may opt for different cutting scores, depending on their tolerance for risk. Cutting scores closer to 0 would render numerically more definitive outcomes, but with an increase in errors. Those wishing to minimize errors further can widen the inconclusive region by separating the cutting scores even more, though there would be an increase in inconclusive outcomes. Appendix B shows the relationship between cutting scores and accuracy with the OSS.

The OSS was not designed to take into consideration the Spot Score Rule. This fact does not preclude the use of the Spot Score Rule, but examiners should be mindful that its effect on accuracy with the OSS has not been published.

The Shortcuts

Readers who have taken the time to wade through the detail of the OSS up to this point are certainly impressed with how involved the procedures are. There are two potential shortcuts that can significantly reduce time and effort requirements of the OSS; one for the measurements, and the other for the computations and scoring. Polygraph examiners who use the Windows version of the Stoelting computer polygraph have available to them an option to have the measurements taken by the polygraph software. This not only saves a considerable amount of time, the precision and reliability of the measurements would probably be better than manual measurements.

Regarding the computational and scoring shortcut, the creator of the OSS maintains a free interactive site on the World Wide Web that will do these operations for examiners. The site is *www.nationalpolygraph consultants.com*. Examiners need only input the measurements described in this paper, and the site will return a score and a suggested decision. An electronic computational spreadsheet is also available without charge by contacting the author by e-mail [ddutton443@aol.com.]

Conclusion

Examiners who are called upon to provide testimony regarding the interpretation of polygraph data are invited to employ the OSS for that purpose. Each of its rules are directly traceable to empirical evidence, a genuine advantage over other methods if one must cite evidence for a given procedure. Moreover, the OSS provides a common method of interpretation that does not require adherence to any of the various schools of thought in the field, thus avoiding one source of disagreement among professionals. It is hoped that it will help improve the validity and reliability of evidentiary polygraphy.

References

Department of Defense Polygraph Institute. (1992). Zone comparison test. Ft. McClellan, Alabama.

Kircher, J.C., & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology*, 73(2), 291-302.

Krapohl, D.J., & McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph*, 28(3), 209-222.

Krapohl, D.J., & Norris, W.F. (2000). An exploratory study of traditional and objective scoring systems with MGQT field cases. *Polygraph*, 29(2).

Timm, H.W. (1982). Analyzing deception from respiration patterns. *Journal of Police Science and Administration*, 10(1), 47-51.

Appendix A. Objective Scoring Worksheet

Chart 1	MEASUREMENTS						RATIOS R/C			SCORES (circle) see table		
	CQ1	**RQ1**	CQ2	**RQ2**	CQ3	**RQ3**	Spot 1	Spot 2	Spot 3	Spot 1	Spot 2	Spot 3
RLL 1										1 resp score only, the stronger.		
RLL 2												
EDA												
BV												
Total												

Chart 2	MEASUREMENTS						RATIOS R/C			SCORES (circle) see table		
	CQ1	**RQ1**	CQ2	**RQ2**	CQ3	**RQ3**	Spot 1	Spot 2	Spot 3	Spot 1	Spot 2	Spot 3
RLL 1										1 resp score only, the stronger.		
RLL 2												
EDA												
BV												
Total												

Chart 3	MEASUREMENTS						RATIOS R/C			SCORES (circle) see table		
	CQ1	**RQ1**	CQ2	**RQ2**	CQ3	**RQ3**	Spot 1	Spot 2	Spot 3	Spot 1	Spot 2	Spot 3
RLL 1										1 resp score only, the stronger.		
RLL 2												
EDA												
BV												
Total												
									Sub totals (all charts)			
											Grand Total	

Scoring Table

RLL	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**
	0.00 to 0.79	0.80 to 0.89	0.90 to 0.96	0.97 to 1.03	1.04 to 1.10	1.11 to 1.25	1.26 to 999
EDA	**-6**	**-4**	**-2**	**0**	**2**	**4**	**6**
	999 to 2.45	2.44 to 1.61	1.60 to 1.21	1.20 to 0.93	0.92 to 0.68	0.67 to 0.44	0.43 to 0.00
BV	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**
	999 to 1.67	1.66 to 1.30	1.29 to 1.06	1.05 to 0.89	0.88 to 0.72	0.71 to 0.54	0.53 to 0.00

Case #_____ Examiner_____ Date_____

Appendix B. Probability estimates for scores of deceptive and nondeceptive cases when using the Objective Scoring System.

Score	Probability of a truthful subject having this score, or lower	Probability of a deceptive subject having this score, or higher
-40	0.01	
-38	0.01	
-36	0.01	
-34	0.01	
-32	0.01	
-30	0.01	
-28	0.01	
-26	0.01	
-24	0.01	
-22	0.01	
-20	0.02	
-18	0.02	
-16	0.02	
-14	0.03	
-12	0.04	
-10	0.05	
-8	0.06	0.20
-6	0.07	0.18
-4	0.09	0.15
-2	0.11	0.13
0	0.13	0.11
2	0.15	0.09
4	0.18	0.08
6	0.21	0.06
8	0.24	0.05
10		0.04
12		0.03
14		0.03
16		0.02
18		0.02
20		0.01
22		0.01
24		0.01
26		0.01
28		0.01
30		0.01
32		0.01
34		0.01
36		0.01
38		0.01
40		0.01

An Exploratory Study of Traditional and Objective Scoring Systems with MGQT Field Cases

Donald J. Krapohl and William F. Norris

Abstract

Three experienced polygraph examiners performed traditional 7-position scoring of 32 confirmed field cases, divided equally between deceptive and nondeceptive, of the Modified General Question Technique format. The deceptive and nondeceptive cases were individually matched against one another for type of crime and confirmation. The same cases were also evaluated using a new Objective Scoring System (OSS) (Krapohl & McManus, 1999). Traditional scoring did significantly better with the deceptive cases than with nondeceptive cases, while the OSS did equally well with both types of cases. Overall error rates were similar for the two scoring systems, but the OSS made a greater proportion of correct decisions. Scoring methodologies were discussed in light of these and other findings.

Key words: 7-position scale, Modified General Question Test, Objective Scoring System, spot scoring, validity, Zone Comparison Technique

In 1999 a new scoring methodology was introduced for evidentiary polygraph examinations (see Krapohl & McManus, 1999 for a complete description), labeled the Objective Scoring System (OSS). The OSS is an adaptation of the traditional 7-position scoring system (Backster, 1963; Swinford, 1999), which assigns whole number values from -3 to +3 to differences in response intensity between relevant and comparison questions. The OSS's principal departure from traditional 7-position scoring is the substitution of subjective estimates of differential responsivity with simplified and objective scoring rules. The method was trained on Zone Comparison Test (ZCT) data garnered from confirmed cases conducted in the field by U.S. Government and law enforcement polygraph examiners. Two cross validations were reported in the Krapohl, et al report, both of which found the new scoring system to make correct decisions of deception and nondeception near 90% or better. At this level of validity, the scoring system applied to ZCT met the standard for evidentiary polygraph examinations set by the American Society for Tests and Materials (ASTM, 1999). The performance of traditional numerical scoring procedures applied to a cross validation data set fell substantially short of the ASTM 90% accuracy criterion.

Extending the earlier work with the OSS, the present pilot study was designed to test OSS efficacy with field data from the Army Modified General Question Test (MGQT). Like the ZCT, the MGQT is a single-issue probable-lie comparison question technique, and is amenable to traditional 7-position scoring. MGQT examinations are routinely used by polygraph examiners in law enforcement and government for criminal investigations. Because of its prevalence in the field, it was considered useful to test the new scoring methodology against MGQT data.

Acknowledgements

The authors are grateful for the assistance of the scorers in this study, to Kay Williams for administrative support, and to Dr. Andrew Dollins and Dr. Stuart Senter for methodological and statistical guidance.. The statements made in this report do not represent the position of the Department of Defense Polygraph Institute or the U.S. Government. For reprints, write to Donald Krapohl, PO Box 10342, Ft. Jackson, SC 29207, or e-mail to dkrapohl@aol.com.

Though the MGQT and ZCT are both probable-lie comparison question tests, the MGQT is dissimilar to the ZCT in three important respects. First, the MGQT typically uses more relevant questions than does the ZCT (4 versus 3). Second, the MGQT contains fewer comparison questions than does the ZCT (2 versus 3). Finally, the question sequence of the MGQT places two relevant questions before the first comparison question is presented. The ZCT has a comparison question placed before each relevant question, including the first relevant question. Because of these structural differences, the generalizability of the OSS, or even traditional scoring, from the ZCT data to the MGQT data is far from certain. The configuration of the MGQT format would seem to make it well suited for detecting deception over the ZCT, given the priority of relevant questions in number and order, but it is equally possible that this advantage is paid for by an increase in false positive outcomes.

One important, and frequently overlooked, feature of probable-lie comparison question tests has to do with the relative magnitudes of responses to the evaluated questions. It has been previously shown that there is a response asymmetry between deceptive and nondeceptive examinees: deceptive examinees tend to respond stronger to relevant questions than nondeceptive examinees respond to comparison questions (Franz, 1988; Kircher, Raskin & Honts, 1994; Krapohl, 1999; Raskin, Kircher, Honts, & Horowitz, 1988). This phenomenon creates a potential source of inefficiency for some scoring systems. Traditional scoring methodology overlays the response asymmetry with symmetrical scoring rules, complicating the task of correctly identifying nondeceptive examinees. The OSS accommodates the asymmetry in its scoring rules, and at least for the ZCT, appears to have a balanced accuracy for both deceptive and nondeceptive examinees. There is no reason to expect that it would not also produce balanced validity and error rates for deceptive and nondeceptive examinees tested with the MGQT, but it has yet to be confirmed.

One other important difference between the OSS and traditional scoring is the inclusion of an additional decision rule for the latter. Most scoring systems have decision rules that address the totals of scores. For many, total scores greater than +5 are deemed No Deception Indicated (NDI), while scores lower than -5 are called Deception Indicated (DI). All scores in between are called Inconclusive or No Opinion (NO). The Spot Score Rule (SSR) is a special decision rule that is applied to the sums of scores for each individual relevant question, and the SSR is always used exclusively with the MGQT testing format. Any relevant question with a total score of -3 or lower would result in a DI outcome, regardless of the total score for all questions. If any question had scores that totaled from -2 to 2, the result would be NO. Only when each relevant question produces a sum of +3 can an NDI decision be made.

With what is known about traditional scoring and the OSS with ZCT, three predictions were made in this exploratory study.

Prediction #1. The better precision of the OSS will result in more correct decisions overall than traditional blind 7-position scoring.

Prediction #2. Because of the symmetrical scoring rules and the SSR, traditional scoring will have significantly more false positive errors than false negative errors.

Prediction #3. The OSS will tend to produce a more equal distribution of errors for deceptive and nondeceptive cases than will traditional scoring.

Method

Polygraph Cases

The design called for an equal number of confirmed deceptive and nondeceptive field cases, matched for type of crime and type of independent confirmation. The entire DoDPI confirmed case database was searched for confirmed field Army MGQT cases. Confirmation was established by confession of the examinee, confession of someone other than the examinee, and physical evidence. The accuracy of the original polygraph decision was not a criterion for inclusion in the database.

In that database, there were only 17 nondeceptive MGQT cases available, thus

limiting the potential size of the study to 34 cases if there were to be an equal proportion of nondeceptive and deceptive cases in the sample. These 17 nondeceptive cases were matched with confirmed deceptive case by type of crime and confirmation. However, after scoring was completed it was learned that one of the selected nondeceptive cases had been listed twice in the database, once as deceptive and again as nondeceptive. We were unable to determine which listing was correct. The scorings for that case were thrown out, along with its matched case, leaving 16 cases each of confirmed deceptive and nondeceptive examinees. The relevant crimes were larceny, sexual assault, and murder.

Given the limited sample size, the efficacy of the two scoring systems were compared only to one another, and there was no attempt to extrapolate the findings to the question of the validity of the MGQT itself. All of the selected cases had three charts, and were conducted on an Axciton computer polygraph (Axciton Systems, Houston, Texas).

While the score sheets of the original testing examiners were not available for the 32 cases in this study, their decisions were. Table 1 lists the accuracy of those polygraph outcomes. The Spot Score Rule was used in the field decisions. In 24 of the 32 cases an independent quality control decision was also available. Since there were no disagreements between the quality control decisions and those of the testing examiners for those cases, those independent decisions are not considered further here.

It is important to note that one of the nondeceptive cases was unlike any of the others in the sample. In that case, #23, the examinee had been called deceptive by the testing examiner, and the decision supported by quality control. The examinee was interrogated regarding the relevant issue, which was the theft of valuable government property. The examinee confessed to another significant theft from the government, but not to the theft under investigation. He was later cleared definitively of the theft for which he underwent polygraph testing. The question arises as to whether this case had a true positive or a false positive polygraph result. It is acknowledged that the utility of the polygraph was demonstrated in this case, since a theft of interest to the government was uncovered after a failed examination about stealing, but that there was the inescapable problem with the specificity of the examination to contend with. Because the two thefts were of different property, and the relevant polygraph questions did not address the true theft the examinee had committed, the case is labeled here as a false positive outcome.

Table 1. Correct, incorrect, and No Opinion decisions of the original testing examiners for the selected confirmed field MGQT cases (n=32).

	Deceptive Cases	Nondeceptive cases	Overall
Correct	13	8	21
Incorrect	0	4	4
No Opinion	3	4	7

Scorers

Three experienced polygraph examiners participated in the study. They had from 5 to 13 years of polygraph field experience. All received initial training at schools accredited by the American Polygraph Association, and were practiced with the 7-position scoring system.

Scoring Rules

The scorers utilized the 7-position scoring rules in accordance with the procedures reported by Swinford (1999). Following the scoring protocol for the MGQT, on the first and second charts the following comparisons were made between the relevant (R) and comparison (C) questions: R3 with C6, R5 with C6, R8 with C10, and R9 with C10.

On the third chart, which is a mixed series, the reactions to each relevant question were assessed against the stronger adjacent comparison question reactions.

Decision Rules

Decisions of NDI require a +3 total for each and every question. If the total score to any one question was -3 or lower, the call is DI, irrespective of the scores for the other questions. If any question had a total of -2 to +2 for any question, the call would be No Opinion. If the total score for each relevant question was +3 or greater, the call is NDI.

Objective Scoring System

The scoring procedures are described in detail in the Krapohl & McManus (1999) report, and will only be briefly discussed here. The diagnostic features for the OSS are respiration line length (RLL), and electrodermal response (EDR) and blood volume (BV) amplitude. The measurements of these features were taken from the raw digitized physiological data using the Extract software package, Version 3 (Johns Hopkins University Applied Physics Laboratory, 1999). The measurements were converted to ratios by dividing the value of the measurements of the relevant question by those of the comparison question (R/C). The relevant and comparison question pairings matched those of the human scorer, listed above under Scoring Rules.

A table was used to assign the scoring value. The ZCT table from Krapohl et al study was used in the present MGQT study, and is reprinted here as Table 2. Once all of the individual scorers were assigned, they were summed. Totals greater than 5 were deemed NDI, those lower than -5 were called DI, and all others were No Opinion. The Spot Score Rule is not applied in the OSS.

Table 2. Ratios for the assignment of scores with the objective scoring system.

	-3	-2	-1	0	1	2	3
RLL	0.00 to 0.79	0.80 to 0.89	0.90 to 0.96	0.97 to 1.03	1.04 to 1.10	1.11 to 1.25	1.26 to 999
	-6	**-4**	**-2**	**0**	**2**	**4**	**6**
EDR	999 to 2.45	2.44 to 1.61	1.60 to 1.21	1.20 to 0.93	0.92 to 0.68	0.67 to 0.44	0.43 to 0.00
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**
BV	999 to 1.67	1.66 to 1.30	1.29 to 1.06	1.05 to 0.89	0.88 to 0.72	0.71 to 0.54	0.53 to 0.00

RLL = Respiration Line Length
EDR = Electrodermal Response
BV = Blood Volume

Results

The three blind scorers correctly classified an average of 14.3 of the 16 deceptive cases. Their collective accuracy for the deceptive cases was 89.4% when No Opinion decisions were counted as errors, and 100% when those outcomes were excluded. Performance with the nondeceptive cases was poor, with correct decisions averaging 1.3 of those 16 cases. Accuracy for nondeceptive cases was 8.3% with No Opinion decisions, and 13.8% without them. Overall accuracy for all cases was 49.0% with No Opinions, and 64.4% without them. (See Appendices A and B for decisions and scores for all scorers.)

Unlike the human evaluators, the OSS made a substantial proportion of errors with deceptive cases. Accuracy with deceptive cases was 75%, and there were no No Opinion results. The OSS appeared to have redeemed itself with the nondeceptive cases, however, correctly identifying 75% of them with no No Opinion results, performing better than even the original examiner. Table 3 shows the decisions of the three scores, the original examiner and the OSS.

Table 3. Number of decisions for deceptive (n=16) and nondeceptive (n=16) cases for the original examiner, three blind scorers, and the Objective Scoring System.

	Deceptive Cases			Nondeceptive Cases			Overall		
	DI	NDI	NO	DI	NDI	NO	Correct	Error	NO
Original Examiner	13	0	3	4	8	4	21	4	7
Scorer 1	12	0	4	7	2	7	14	7	11
Scorer 2	16	0	0	10	2	4	18	10	4
Scorer 3	15	0	1	8	0	8	15	8	9
OSS	12	4	0	4	12	0	24	8	0

DI = Deception Indicated
NDI = No Deception Indicated
NO = No Opinion
OSS = objective scoring system

Proportions of agreement were calculated for all pairs of decisions, and the decisions against ground truth. Table 4 lists those proportions. Not all were significantly better than chance at the .05 level.

Table 4. Proportions of decision agreement between pairs of scorers, and between scorers and ground truth.

	Scorer 1	Scorer 2	Scorer 3	OSS	Ground Truth
Original Examiner	0.63*	0.56	0.53	0.59*	0.66
Scorer 1		0.72*	0.75*	0.53	0.44
Scorer 2			0.78*	0.56	0.56
Scorer 3				0.50	0.47
OSS					0.75*

OSS = objective scoring system
* denotes those proportional values significantly different from chance ($p<.05$).

Prediction #1. The better precision of the OSS will result in more correct decisions overall than traditional blind 7-position scoring.

Of the 32 cases, the average proportion of correct decisions for the blind scorers was 0.49, compared to 0.75 for the OSS. The test

of proportional differences was significant ($z=2.14$, $p<.05$). Prediction #1 was supported.

The average percentage of errors for the three blind scorers was 51.0% when NOs were considered, and 34.7% when they were excluded. Average errors for OSS was 25.0% for both the with-NOs and without-NOs conditions, since it did not render any NOs with this sample. In the with-NOs condition, the average proportion of error for the blind scorers was significantly greater than the OSS error rate ($z=2.14$, $p<.05$) but not in the without-NOs condition ($z=0.85$, ns). Therefore, the OSS had significantly fewer errors than the average blind scorer only in the with-NOs condition.

Prediction #2. Because of the symmetrical scoring rules and the Spot Score Rule, traditional scoring will have significantly more false positive errors than false negative errors.

None of the human scorers made a false negative error with the 16 deceptive cases, and they produced an average of 8.3 false positive errors with the nondeceptive cases. The proportional difference was significant ($z=3.27$, $p<.05$), and prediction #2 is supported. To assess the contribution of the Spot Score Rule on the unbalanced error rates, a post hoc analysis of decision accuracy was made of the human scorings where the Spot Score Rule was removed. Table 5 shows the accuracies when only the grand sum was used to render decisions, with cutting scores of +/-6.

Table 5. Decisions for 3 blind scorers of MGQT data without the Spot Score Rule.

	Deceptive Cases			Nondeceptive Cases			Overall		
	DI	NDI	NO	DI	NDI	NO	Correct	Error	NO
Scorer 1	8	3	5	4	9	3	17	7	8
Scorer 2	12	2	2	4	6	6	18	6	8
Scorer 3	9	0	7	4	8	4	17	4	11
Average	9.67	1.67	4.67	4.00	7.67	4.33	17.33	5.67	8.33

Errors for the deceptive cases were still lower than those for the nondeceptive cases (1.67 versus 4.00), but the values were no longer significantly different ($z=1.08$, ns). The SSR with these data permitted an increase in true positive decisions from 60.4% to 91.7% over not using the SSR. The cost for the SSR was a reduction in true negatives from 47.9% to 8.3%. Combining all decisions, there was an increase in accuracy from 49.0% to 54.2% when the SSR was ignored, but this difference was not significant ($z=0.42$, ns). The increase in accuracy attributable to the SSR for the deceptive cases was 31.3 percentage points, and the decrease for the nondeceptive was 39.6 points. This approximately equal trade-off in accuracy for the deceptive and nondeceptive for which the SSR is responsible is consistent with the findings with laboratory data (Krapohl, 1998). The generic effect of the SSR is the subject of another study currently underway.

Prediction #3. The OSS will tend to produce a more equal distribution of errors for deceptive and nondeceptive cases than will traditional scoring.

False positive and false negative errors were equal at 4 each for the OSS, compared to an average of 8.3 and 0.0 for the blind scorers. The difference in error rates for the two types of cases (deceptive and nondeceptive) for the blind scorers was significant, as was discussed under Prediction #2 above. There was no imbalance in error rates for the OSS. Prediction #3 was supported.

Discussion

The present restricted sample size precluded a comparison of the validity of the MGQT to that of the ZCT with the traditional and objective scoring systems. However, we are able to make some preliminary statements

of the scoring methods under review. First, we have added to the growing body of findings that traditional numerical scoring of the CQT is more sensitive to deception than to truthfulness. Second, it would appear that using asymmetrical scoring rules, such as those of the OSS, not only can the shift toward deceptive decisions be overcome, but the proportion of correct decisions can also be increased. This was found both in the earlier Krapohl et al ZCT study (1999), and the present MGQT study.

One of the principal criticisms leveled against polygraphy in general is that, while it catches liars very well, it cannot detect truthfulness with equal accuracy (Lykken, 1998). There are field studies that suggest that these criticisms can be supported to some extent (Horvath, 1977; Kircher, Raskin & Honts, 1994; Patrick & Iacono, 1989; Raskin & Hare, 1978). Part of the argument is based on a contested accounting method: critics of polygraphy count No Opinions as errors, while proponents do not. Notwithstanding the issue as to whether a No Opinion is a decision, the proportion of true negative outcomes is almost always lower than true positive outcomes in blind scorings of field cases. Therefore, regardless of the tallying method, it is fairly well accepted that conventional polygraphy finds liars better than it finds truthtellers.

Why might this be so? For purposes of discussion, let us consider the testing technique separate from the chart analysis method. The present study, along with the earlier ZCT study, would suggest that the problem is not entirely that the testing techniques are biased (though there remains work to be done there, too), but that the traditional methods of chart analysis are built on incorrect assumptions, and that they are responsible for at least part of the trend toward diminished sensitivity to truthfulness. One such assumption is that nondeceptive examinees react in precisely an opposite pattern from what is seen with deceptive examinees, vis a vis responses to relevant and comparison questions. Such a belief is unsupportable by any study with field data found by the present authors, and the evidence is beginning to accumulate that the assumption is wrong.

Another unproven assumption is that the inclusion of localized decision rules, such as the Spot Score Rule, increases accuracy (see Light, 1999). The available evidence tentatively suggests that the SSR does little or nothing to increase accuracy, as it robs validity from the nondeceptive cases to pay for the validity of the deceptive cases. This is not to condemn the SSR entirely. The SSR would, indeed, help increase the number of correct decisions if the base rate of deceptive examinees is very high. This is because the SSR makes the examination sensitive to what it would face most often. For example, if the base rate of deception in a tested population were 90%, a polygrapher's overall hit rate would be better with scoring rules that improve the chances of detecting deception, even at a loss of detection of nondeception. Conversely, low base rates of deceptive examinees, or when the cost of a false positive error is high, argue against the SSR, and users should be circumspect about its application across the board. When considering the SSR in single-issue examinations it is very important to be mindful of the expenses associated with errors. Depending on the costs of false positives (i.e. merely posttest questioning versus a criminal conviction) the SSR may have a place among scoring procedures under the high deception base rate condition. However, when a DI decision would levy a significant penalty on the examinee, and the base rate of deceptiveness is balanced or low, the SSR in single-issue examinations cannot be justified under the current understanding.

The present project was a preliminary assessment of the relative accuracy of traditional and objective scoring. Readers are cautioned that the modest sample size would not allow generalization of the validity of the MGQT found with this sample. Nevertheless, the present results are consistent with earlier findings with ZCT data, that the OSS is balanced, and that it performs at least as well as human scorers overall. Practitioners are encouraged to test it further with their confirmed single-issue confirmed MGQT cases.

References

Backster, C. (1963). The Backster chart and reliability rating method. *Law and Order*, 11, 63-64.

Franz, M.L. Relative contributions of physiological recordings to detect deception. Technical Report contract # MDA904-88-M-6612, Argenbright Polygraph, Inc., Atlanta, GA

Horvath, F.S. (1977). The effect of selected variables on interpretation of polygraph records. *Journal of Applied Psychology*, 62, 127-136.

Kircher, J.C., Raskin, D.C., & Honts, C.R. (1994). Generalizability of statistical classifiers for the detection of deception. *Psychophysiology*, 31(Supplement 1), S11 (Abstract).

Krapohl, D.J. (1998). A comparison of 3- and 7-position scoring scales with laboratory data. *Polygraph*, 27(3), p 210 – 288.

Krapohl, D. (1999). Short report: Proposed method for scoring electrodermal responses. *Polygraph*, 28(1), 82-84.

Krapohl, D.J., & McManus, B. (1999). On objective method for manually scoring polygraph data. *Polygraph*, 28(3), 209-222.

Light, G. (1999). Numerical evaluation of the Army Zone Comparison Test. *Polygraph*, 28(1), 37-45.

Lykken, D.T. (1998). A tremor in the blood: Uses and abuses of the lie detector. New York: Plenum Trade.

Patrick, C.J. & Iacono, W.G. (1989). Psychopathy, threat and polygraph test accuracy. *Journal of Applied Psychology*, 74(2), 347-355.

Raskin, D.C., & Hare, R.D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology*, 15, 126-136.

Raskin, D.C., Kircher, J.C., Honts, C.R., & Horowitz, S.W. (1988). A study of the validity of polygraph examinations in criminal investigations: Final report to the National Institute of Justice. Grant No. 85-IG-CX-0040.

Swinford, J. (1999). Manually scoring polygraph charts utilizing the seven-position numerical analysis scale at the Department of Defense Polygraph Institute. *Polygraph*, 28(1), 10-27.

Appendix A. Decisions and ground truth by case

Case	Ground Truth	Decisions				
		Orig Call	Scorer 1	Scorer 2	Scorer 3	OSS
1	DELETED					
2	Nondeceptive	NDI	No Opinion	No Opinion	No Opinion	NDI
3	Nondeceptive	NDI	NDI	NDI	No Opinion	NDI
4	Deceptive	No Opinion	No Opinion	DI	DI	NDI
5	Nondeceptive	No Opinion	DI	DI	DI	DI
6	Deceptive	No Opinion	DI	DI	DI	DI
7	Nondeceptive	NDI	No Opinion	No Opinion	No Opinion	NDI
8	Deceptive	DI	DI	DI	DI	DI
9	Nondeceptive	NDI	No Opinion	No Opinion	DI	NDI
10	Nondeceptive	NDI	No Opinion	NDI	No Opinion	NDI
11	Nondeceptive	No Opinion	NDI	DI	No Opinion	NDI
12	Nondeceptive	No Opinion	DI	DI	No Opinion	NDI
13	Nondeceptive	DI	No Opinion	No Opinion	No Opinion	NDI
14	Nondeceptive	No Opinion	No Opinion	DI	DI	NDI
15	Deceptive	DI	DI	DI	DI	DI
16	Deceptive	DI	DI	DI	DI	DI
17	Deceptive	DI	DI	DI	DI	DI
18	Nondeceptive	DI	DI	DI	DI	NDI
19	Deceptive	No Opinion	DI	DI	DI	DI
20	Nondeceptive	NDI	NDI	No Opinion	No Opinion	NDI
21	Deceptive	DI	DI	DI	DI	DI
22	Deceptive	DI	DI	DI	DI	DI
23	Nondeceptive	DI	DI	DI	DI	DI
24	Deceptive	DI	No Opinion	DI	No Opinion	NDI
25	Deceptive	DI	DI	DI	DI	DI
26	Deceptive	DI	DI	DI	DI	NDI
27	Deceptive	DI	No Opinion	DI	DI	NDI
28	Nondeceptive	NDI	DI	DI	DI	DI
29	Deceptive	DI	DI	DI	DI	DI
30	Nondeceptive	DI	DI	DI	DI	DI
31	Deceptive	DI	DI	DI	DI	DI
32	Deceptive	DI	DI	DI	DI	DI
33	DELETED					
34	Nondeceptive	NDI	DI	DI	DI	NDI

DI = Deception Indicated
NDI = No Deception Indicated
NO = No Opinion

Case	Ground Truth	Scorer 1					Scorer 2					Scorer 3					OSS				
		R3	R5	R8	R9	Total	R3	R5	R8	R9	Total	R3	R5	R8	R9	Total	R3	R5	R8	R9	Total
1	DELETED																				
2	Nondeceptive	5	2	9	10	26	1	1	8	8	18	0	-1	5	5	9	7	5	32	26	70
3	Nondeceptive	11	18	12	12	53	13	6	14	17	50	9	2	12	9	32	30	22	27	23	102
4	Deceptive	4	1	5	2	12	6	-1	5	-4	6	2	-3	1	-2	-2	11	-7	11	0	15
5	Nondeceptive	-7	-8	-3	-1	-19	-6	-5	-3	-1	-15	-5	-6	-1	-2	-14	-24	-22	-1	-2	-49
6	Deceptive	-3	-5	-3	-1	-12	-5	-3	-2	-2	-12	-4	-2	-3	1	-8	-16	-16	0	-2	-34
7	Nondeceptive	0	0	5	7	12	-2	-2	3	4	3	-2	-1	0	6	3	2	5	4	20	31
8	Deceptive	-8	-7	3	0	-12	-9	-9	1	-4	-21	-7	-5	0	-3	-15	-9	-19	5	-3	-26
9	Nondeceptive	-1	-2	-1	0	-4	3	1	-1	1	4	3	1	-3	1	2	-5	-1	4	9	7
10	Nondeceptive	0	3	3	2	8	3	4	9	6	22	0	2	5	4	11	7	14	18	10	49
11	Nondeceptive	4	4	7	3	18	4	3	4	-6	5	2	1	5	-2	6	4	11	18	4	37
12	Nondeceptive	-3	-4	9	3	5	-6	-7	7	2	-4	1	0	3	2	6	-4	-6	24	2	16
13	Nondeceptive	-1	0	6	4	9	3	1	13	6	23	-2	-1	8	4	9	-2	-7	21	19	31
14	Nondeceptive	0	3	14	6	23	-3	-3	10	9	13	-6	-1	10	5	8	-10	-3	25	16	28
15	Deceptive	-4	0	2	-3	-5	-8	1	3	-3	-7	-3	-1	0	-3	-7	-21	2	5	-4	-18
16	Deceptive	-4	-9	-1	-5	-19	-6	-8	1	-2	-15	-6	-4	2	-1	-9	-11	-16	6	-5	-26
17	Deceptive	-5	-4	3	5	-1	-6	-8	-5	-1	-20	-9	-10	-6	-1	-26	-17	-22	-12	-10	-61
18	Nondeceptive	-7	2	-2	-1	-8	-8	2	1	-2	-7	-5	0	-1	0	-6	-9	17	0	3	11
19	Deceptive	-1	-3	0	2	-2	1	-5	-4	-1	-9	-1	-3	0	0	-4	4	-5	-9	2	-8
20	Nondeceptive	4	10	9	4	27	6	9	10	1	26	3	5	8	2	18	8	16	26	7	57
21	Deceptive	-5	-4	-7	-8	-24	-5	-5	-7	-8	-25	-5	-6	-9	-9	-29	-13	-7	-15	-32	-67
22	Deceptive	-9	0	-1	1	-9	-11	-4	-4	-5	-24	-7	-3	0	-3	-13	-22	-9	-19	-12	-62
23	Nondeceptive	-6	-9	-4	0	-19	-2	-8	-6	-4	-20	-5	-5	-3	-2	-15	-5	-19	4	-2	-22
24	Deceptive	4	4	4	1	13	-7	1	2	1	-3	1	0	0	0	1	1	0	1	5	7
25	Deceptive	-7	-12	-6	0	-25	-11	-8	-10	-5	-34	-9	-9	-6	0	-24	-12	-25	-20	-4	-61
26	Deceptive	-2	-4	-2	1	-7	-5	-5	1	1	-8	-1	-2	-3	2	-4	-2	-3	10	5	10
27	Deceptive	-1	2	4	3	8	-7	1	11	2	7	0	-3	2	1	0	-1	15	21	14	49
28	Nondeceptive	-6	-5	-2	-2	-15	-13	-9	-4	-4	-30	-5	-4	-3	-4	-16	-5	-21	1	-13	-38
29	Deceptive	-5	0	2	1	-2	-4	0	3	2	1	-3	1	1	-1	-2	-8	-13	-4	-17	-42
30	Nondeceptive	-1	-5	-2	6	-2	0	-6	-2	4	-4	-1	-3	-1	3	-2	-9	-18	3	10	-14
31	Deceptive	-3	4	3	1	5	-4	-2	-8	-9	-23	-2	2	-3	-2	-5	-21	-13	-14	-8	-56
32	Deceptive	-2	-7	-4	-3	-16	-2	-7	-3	-5	-17	-2	-5	-3	-5	-15	-9	-7	4	-16	-28
33	DELETED																				
34	Nondeceptive	1	-5	4	7	7	1	-7	4	7	5	-1	-6	-1	3	-5	5	-9	5	16	17

Short Report

A Comparison of 3- and 7-Position Scoring Scales with Field Examinations

Esther M. Harwell

This report outlines a replication of a methodology used by Krapohl (1998) in which he compared 3- and 7-position scoring scales with laboratory data. The goal of Krapohl's study was to determine which cutting scores applied to the 3-position scoring system would yield equivalent outcomes to those of the 7-position scoring system when +/-6 cutting scores were used. Spot scores were not considered in that effort, nor in the present project. His study, utilizing laboratory data, determined that the 3-position system thresholds of +/- 3, +/- 4 and +/-5 produced proportions of outcomes that were not significantly different from those rendered by the orthodox +/- 6 thresholds of the 7-position scale. Of those, +/- 4 produced the least total variations from the decisions reached by the 7-position scores. This study was conducted to determine if Krapohl's findings using laboratory data would generalize to field examination data.

Two previous studies pertaining to the 3-position and 7-position scales have been conducted; one by Capps and Ansley (1992), and another by Van Herk (1991). Both found cutting scores for the 3-position scoring system that closely approximated the outcomes from the 7-position scoring system. Capps et al concluded that +/-3 provided the best fit, while Van Herk arrived at +/-4. Each used different methodologies to arrive at these conclusions, and despite very similar findings, neither had assessed the full range of cutting scores. The present study is the first to include a systematic analysis of all cutting scores from +/-1 to +/-6 for the 3-position scale using field data.

In the current project, 88 sets of physiological detection of deception (PDD) recordings were utilized. All were field examinations, conducted with the Zone Comparison Technique (ZCT) question format as taught by the Department of Defense Polygraph Institute (DoDPI). The ZCT examinations were those used by Blackwell (1999) in another scoring study. Blackwell had three experienced PDD examiners score 100 sets of confirmed ZCT examinations, all of which had resulted from actual criminal examinations. Of those 100 examinations, 88 were used here. Twelve of the examinations were eliminated because they were conducted using a two-question ZCT format. To replicate Krapohl's design, only three-question format ZCT examinations were considered, reducing the 100 examinations to 88. Of the 88 examinations, 60 were confirmed deceptive, and 28 were confirmed nondeceptive.

Consistent with Krapohl's methodology, this study utilized the values generated by the examiners using the 7-position scoring scale, converting them to a 3-position scale (-1, 0, +1), by collapsing the +2 and +3 values to a +1, and the -2 and -3 values to a -1. Then all values were summed across all questions and all tests, to render a single total value for each examination.

Results

Using the standard 7-position scale with cutting scores of +/-6, the three scorers were correct 68.5% of the time. They were incorrect 2.6% of the time, and in 28.7% of the cases they produced NO results. These results

The author is an active duty U.S. Army Criminal Investigation Division Command (USACIDC) polygraph examiner, serving as an instructor at the Department of Defense Polygraph Institute (DoDPI). She is also a member of the American Polygraph Association. The conclusions expressed in this paper do not necessarily represent the views of the US Government or the American Polygraph Association.

Acknowledgements: The author is grateful to Mr. Donald Krapohl for technical guidance.

were based on 88 cases, evaluated by three examiners, for a total of 264 decisions. These results are further broken down by ground truth. The three sets of scorings for the 28 confirmed nondeceptive examinations yielded 84 decisions (48 correct, 4 incorrect, and 32 No Opinion (NO) results). The three sets of scorings of the 60 confirmed deceptive examinations produced 180 decisions (133 correct, 3 incorrect, and 44 NO results). The overall proportion of correct decisions was significantly greater than chance ($z=4.34$, $p<.01$). A total of 71.2% of the results were conclusive, and excluding NOs, the scorers averaged 96.3% correct decisions. Table 1 displays the relative accuracy of the 7-position scale, along with the 3-position scale at various symmetrical cutting scores.

Table 1
Number of correct, No Opinion, and incorrect decisions for 3 scorers of 88 sets of polygraph charts by scoring system and cutting scores (n=264 decisions per system).

	Decisions		
Method & Cutting Score	Correct	No Opinion	Error
7-position (+/-6)	**181**	**76**	**7**
3-position (+/-6)	151	108	5
3-position (+/-5)	166	91	7
3-position (+/-4)	**177**	**78**	**9**
3-position (+/-3)	198	49	17
3-position (+/-2)	217	24	23
3-position (+/-1)	224	12	28

The performance of the 3-position scale was a function of the thresholds. In order to find the best match for outcomes of the 7-position scoring system, a series of goodness of fit chi-square analyses were conducted for all cutting scores between +/-1 to +/-6 for the 3-position scoring data. Of those cutting scores, only two were not significantly different from the 7-position outcomes at the .05 level: +/-4 ($\div^2=0.32$, $p>.05$) and +/-5 ($\div^2= 2.00$, $p>.05$). Of those two sets of cutting scores, +/-4 produced proportions of accuracies the least dissimilar from the 7-position scoring system with its +/-6 cutting scores. Van Herk's study, looking only at cutting scores of +/-4 through +/-6 for the 3- position scale, also concluded that +/-4 produced outcomes closest to those of the 7-position scale. Capps et al did not conduct a systematic analysis of cutting scores, but did find that switching from the 7-position scale to the 3-position scale without changing the cutting scores from +/-6 led to an excessively high NO rate. The present data would also suggest an exceptionally high NO rate for the 3-position scale with cutting scores of +/-6.

Table 2 places the performance of the 7-position scoring at +/-6 and 3-position scoring systems at +/-4 cutting scores. Table 3 depicts the proportion of agreement among the three scorers employing the 3-position scale, utilizing three charts with +/-4 cutting scores.

Table 2
Number of correct, incorrect, and No Opinion results from 7-position at +/-6 cutting scores, and 3-position scoring at +/-4 cutting scores.

Ground Truth	Polygraph Decision	7-Position	3-Position
Nondeceptive	No Deception Indicated	48	53
Nondeceptive	No Opinion	32	25
Nondeceptive	Deception Indicated	4	6
Deceptive	No Deception Indicated	3	3
Deceptive	No Opinion	44	53
Deceptive	Deception Indicated	133	124

Table 3
Proportions of agreement on polygraph decisions among three scorers and ground truth employing the 3-position scale, three charts, and +/-4 cutting scores.

	Scorer 2	Scorer 3	Ground Truth
Scorer 1	0.65	0.68	0.64
Scorer 2		0.70	0.74
Scorer 3			0.65

Discussion

The purpose of this research was to replicate Krapohl's methodology, but with field data, to determine the cutting scores for the 3-position scale that would match the proportions of outcomes resulting from the 7-position scale. The present data also found that +/-4 was the best of all symmetrical cutting scores from +/-1 to +/-6, as did Krapohl. These converging lines of evidence would suggest that those examiners utilizing the 3-position scoring system with 3-question 3-chart ZCT examinations could apply these lower cutting scores without significantly affecting accuracy over the 7-position scoring. Practitioners should be mindful that cutting scores, whether with the 3- or 7-position scoring scales, will affect accuracy, and while the conventional +/-6 thresholds were used for comparison here, this does not imply that these are optimal. More work remains to determine which cutoff scores provide the best accuracy.

References

Blackwell, J.N. (1999). PolyScore 3.3 and Psychophysiological Detection of Deception Examiner Rates of Accuracy When Scoring Examinations from Actual Criminal Investigations. Reprinted in *Polygraph*, 28(2), 149-166.

Capps, M.H. & Ansley, N. (1992). Comparison of two scoring scales. *Polygraph*, 21(1), 39-43.

Krapohl, D.J. (1998). A comparison of 3- and 7-position scoring scales with laboratory data. *Polygraph*, 27(3), 210-218.

Van Herk, M. (1991). Numerical evaluation: Seven point scale +/-6 and possible alternatives: A discussion. *Polygraph*, 20(2), 70-79.

FLASH POINT
The American Mass Murderer

A book by

Michael D. Kelleher

Praeger Publishers, 88 Post Road West, Westport, CT 06881
Price $55.00, ISBN 0-275-9592925-2, 224 pages, 1997.
Order by toll free number 1-800-225-5000

Review by Dan V. Weatherman

The author, Michael D. Kelleher "specializes in strategic management, human resource management, staff education, threat assessment, and management crisis resolution for organizations in the private and public sectors. He has written *Profiling the Lethal Employee* (1997) and *New Arenas for Violence* (1996), both published by Praeger".

The author points out in his book that between 1976 and 1991, there were approximately 350 incidents of mass murder in the United States in which the number of victims exceeded three. He differentiates between mass murder and crimes committed by serial killers. According to the author, mass murder is an act where multiple individuals (at least three) are intentionally killed by a perpetrator in a single incident. Whereas, a serial killer will kill a number of individuals over a protracted period of time ranging from a few months to many years. He points out that mass murder does not catch the media's attention like that of serial killings because it (mass murder) is a one-time event that is horrible when it happens, but is soon forgotten. The goal of the book appears to be an attempt to understand the psychological processes that lead to an individual committing mass murder. The author takes an untraditional approach and identifies seven broad categories of mass murder. The categories are *"perverted love", "politics and hate", "revenge", "sexual homicide", "mass murder by execution", "sane and insane",* and *"unexplained"*. The author has picked approximately fifty-three different cases of mass murder that he has placed in the seven categories. In each case he provides a brief synopsis of the perpetrator, the incident itself, the weapons used, and the number of people killed. He also attempts to identify the "triggering mechanism" and motivation that caused the perpetrator to initiate the mass murder rampage and the disposition of the killer after the act.

If you have an interest in reading brief synopses of mass murder cases this is the book for you. However, if you are looking for a book delving into the psychology of mass murder, this book would not serve that purpose.

"Detektor lzhi" (Lie Detector)

A book by Varlamov, VA

Krasnodar: Sovetskaya Kuban'. - 1998 (in Russian).

Review by Vitaliy I. Egorov

The appearance of this textbook is a remarkable event not only for Russia but surely for all countries of the former Soviet Union where polygraphy is accepted. Unfortunately, in Russian there are not any of its own, or even translated fundamental books dealing with the science of forensic psychophysiology.

The author of the book is a Doctor of Biological Sciences and the founder of polygraphy in the USSR. In the mid-50s, with his young colleague Dr. A. Suchov, they developed their first polygraph that met international criteria. Being psychiatrists, they primarily used their "novelty" in forensic psychiatric investigation trying to differentiate true psychopathology from simulative behaviors in criminal suspects. But these findings were only experimental. Moreover, Varlamov's attempts to legislate polygraphy brought about a strong political pressure from numerous governmental institutions of the former Soviet Union. Only in 1994 did polygraphy obtain its official status as an effective tool for fighting crime.

This book was written with much emotion. It is not surprising because, in a remarkably short time, Russia has produced about 300-500 polygraph examiners, and according to the APA, this number places Russia a second place in the world only after the USA! However, it is a case where quantity doesn't mean quality. Most polygraph examiners have been trained in Moscow and Krasnodar and their qualification and level of education are far from the internationally recognized criteria for polygraph science.

The book consists of 12 chapters. They discuss the history of polygraph where the author has mentioned the contributions of outstanding Russian psychologists - A.R. Luria and A.N. Leontiev. It also discusses such actual problems of polygraph as its psychophysiological bases, methods of recording and measurement of the data, and the use of PDD in different criminal investigations. Of special interest is a chapter that deals with such important problems of modern forensic psychophysiology as test construction. The author notes that, while recognizing the experience of the USA and other countries, Russia developed its own way in polygraphy that, according to author's point of view, reflects the unique mentality of Russians, and those of criminal elements of this country. In this case let me note that American psychologist Paul Ekman who, after his visit to the Soviet Union, remarked that he saw the most deceptive country that he could visit.

As the first textbook on polygraph in Russia, it has some shortcomings. The author's critique of existing techniques sounds more affectively colored than constructive. Undoubtedly, we have differences in our behaviors, temperament, and probably in intelligence and cognitive "algorithms", but they are not so important for the PDD examination. We still measure mainly the responses of the autonomous nervous system that is evolutionary close to the reptilian brain. The author argues that such adopted techniques as ZCT, MGQT, and I/R are relatively ineffective in the post-Soviet conditions. The psychology of Russians, wrote Varlamov, is principally different. Based on his own large experience and experience of the Krasnodar school of polygraph training, Varlamov supports the use of a limited number of techniques that are close to the POT. We can accept this affectively colored conclusion or not, but must accept the actual fact – polygraph still works in Russia and works effectively.

But, even acknowledging the shortcomings above, I think that the first textbook on polygraph science in Russia is an extraordinary event, and will be a helpful tool in the development of scientific and practical bases of polygraphy, not only in Russia but in all countries of the former USSR.