

VOLUME 21

1992

NUMBER 4

Contents Numerical Scoring of Polygraph Charts: What Examiners Really Do 264 Michael H. Capps and Norman Ansley Anomalies: The Contributions of the Cardio, Pneumo, and Electrodermal Measures Towards a Valid Conclusion 321 Michael H. Capps and Norman Ansley Strong Control versus Weak Control 341 Michael H. Capps and Norman Ansley Historical Note: Mackenzie's Early Polygraph Instrument 349

NUMERICAL SCORING OF POLYGRAPH CHARTS: WHAT EXAMINERS REALLY DO

Michael H. Capps and Norman Ansley

We know something about what examiners are taught to do in reading charts of the type that may be subjected to numerical We also know something about global, rank order, analysis. statistical, computerized, and non-standard methods. There is a considerable body of literature on the results of independent numerical scoring of sets of charts from real crimes and mock crimes where the truth or deception is definitely known. Although numerical scoring is used with most control question tests, there is a difference in the way people apply numerical scoring, partly because of diverse rules, and partly from the fact that arbitrary scoring rules cannot be uniformly applied to the many variations that appear in physiological tracings. There is some literature that suggests that for those test formats that lend themselves to numerical scoring, the use of numerical scoring produces a higher degree of accuracy than global methods, but there are also some contrary findings. The history of reading the physiological recordings has not been one of unified rigorous scientific inquiry. Rather, it appears that most of the current concepts came from the codified observations of experienced polygraph examiners.

What this study sought was information on what criteria polygraph examiners currently apply when they analyze sets of control question polygraph charts. We wanted to know what they did when they were correct, what they did when they were wrong, and what was the difference. We were interested in what difference existed when examiners scored against the weakest control question response compared to the strongest control question response, a fundamental difference between two widely used scoring methods with zone comparison charts. The Backster system scores against the weakest, the DoDPI and other systems score against the strongest, and at one time the Utah system just scored the control preceding the relevant (Raskin, 1979; Weaver, 1985), but that may have changed. We wondered if some deception criteria are being taught that are never used, and that related reactions may rarely or never occur.

The senior author is a past president of the APA and Life Member who has been a regular contributor to the journal. The junior author is a Life Member of the APA and the Editor-in-Chief of APA Publications. For reprints write to P.O. Box 794, Severna Park, MD 21146.

<u>History of Chart Interpretation</u>

The pioneers in polygraph research and practice were men who tried a variety of methods, and kept records or wrote about what worked in the laboratory and what worked in the field, and they arqued about the difference (Trovillo, 1953; Marcuse, 1953). They tried many test formats and a multitude of measures. Through speculation, trial, and error, three physiological measures emerged as being consistently useful: respiration, cardiovascular activity, and electrodermal response. Equipment to record these measures on polygraph instruments was available at the turn of the century (MacDonald, 1905). Cesare Lombroso was probably the first to use a physiological measure for lie detection. Recording vascular volume on a smoked drum from Mosso's hydrosphygmograph, he successfully detected deception and verified truth in a number of criminal cases (Lombroso, 1895, 1911). He may have been the first to use a peak of tension test when he made recordings while showing a suspect a series of photographs of children, one being the victim of sexual assault. When the suspect did not react more to the victim's picture than the pictures of others, Lombroso concluded that since the suspect didn't know what the victim looked like, he was not the perpetrator. His conclusion was subsequently verified (Lombroso-Ferrera, 1911). Lombroso also tells us that he measured blood pressure changes in multimeters, but we don't know if these were changes on his graph or graph changes calibrated to the Korotkoff method of determining arterial pressure changes recorded in millimeters of mercury (mm.Hg). Lombroso (1911), in deciding that a prisoner was guilty of one offense, and innocent of another, said:

"An investigation with the hydrosphygmograph confirmed me in my observation of his [Bersone Peirre] great insensibility to pain, which did not change the sphygmographic lines. The same apathy persisted when he was spoken to of the robbery on the railroad, while there was an enormous depression - a fall of 14 mm - when the Torelli theft was mentioned. I concluded, therefore, that he had no part in the railway robbery, but that he had certainly participated in the Torelli affair; and my conclusions were completely verified."

Here we have any early report of a cardio reaction, verified as deceptive, in which Lombroso compared the reaction to the mention of one offense to the reaction to mention of another offense, and also to the lack of reaction to pain from electric shock. The shock was administered by an adjustable Rhumkorff coil, prior to the test about the crimes. The elements of a control question test were present, if you consider the shock response as a control. The comparison of responses to two similar crimes suggests the later use of guilt complex questions.

In a chapter on the "Traces of Emotions" in Hugo Munsterberg's 1908 book <u>On the Witness Stand</u>, he suggests the possibility of detecting deception, and chooses three instruments. (As Chairman of the Department of Psychology at Harvard, he later supervised the doctoral research of William M. Marston.) What is remarkable about the 1908 book is that Munsterberg chose the three measures that were to become the standard components for detection of deception for the next 80 to 90 years. Note, too, that even at that time Munsterberg was suggesting the best test was for guilty knowledge. In the pneumograph, Munsterberg said it was easy to measure:

- a. the curve
- b. change in length
- c. change in height
- d. change in angle
- e. regularity of the wave:
 - 1) rapidity
 - 2) rhythm
 - 3) distribution
 - 4) pauses and strength of the breathing

Munsterberg noted that pleasure makes the respiration weaker and quicker; displeasure, stronger and slower; excitement makes it stronger and quicker; acquiescence, weaker and slower.

For the galvanometer, Munsterberg said that when the brain is excited or there is any emotional disturbance, this influences the sweat glands, and their functioning alters the electrical conditions.

For the wrist sphygmograph, Munsterberg said it recorded:

- a. height of every pulse
- b. length of every pulse
- c. form of every pulse
- d. a different pulse for inhalation than exhalation

Munsterberg said pleasure heightens and retards the pulse, displeasure weakens and accelerates it, excitement makes the pulse stronger and quicker, and acquiescence, weaker and slower. For the hydrosphygmograph he added that changes in the volume of blood supply could be measured.

Munsterberg observed that they knew "too little about the evident individual differences to make, for instance, a breathing

and pulse curve a basis for a legal condemnation or acquittal." At that point, they could not discriminate the nervous innocent man from the guilty. He said the real use of the experimental emotion method was then "probably confined to those cases in which it was to be found whether a suspected person knew anything about a certain place or thing. Thus, if a new name is brought in, the innocent who never heard the name before will not be more excited if he hears that among a dozen others; the criminal, who knows the name as that of a witness to a crime, will show the emotional symptoms." Thoughtfully, Munsterberg added:

"And yet, it may be rash to propose narrow limits for the practical use, as the rapid progress of experimental criminopsychology may solve tomorrow those difficulties which seem to stand in the way today."

Following the earlier suggestion of Wertheimer, and observations of Mosso, Vittorio Benussi conducted a research project involving respiration and the effect of lying (Herbold-Wooten, 1982). He found that the inspiration/expiration rate was different when someone lied. His 1914 article gave, as illustrations, 16 chart segments with the matching I/E ratio These were the first illustrations of charts for computations. interpretations (Benussi, 1914, 1975, Ruckmick, 1936). More illustrations and tables of I/E ratios appears in Felecky's (1916) More work on emotion and respiration. However, Burtt, using different recording equipment, was unable to replicate Benussi's findings with respect to ratios and deception (Burtt, 1921). Ιn retrospect, Benussi's basic concept was probably correct, as evidenced the more recent work by Howard Timm (1982) with respiratory recording line length. Like Timm's method, Benussi's method did not disclose the full diagnostic value of the respiratory tracing. However, the I/E ratio remains as deception criteria in many modern lists.

Galvanometers were also in use at the turn of the century (MacDonald, 1905) having been developed by Fere (1888). They were used in experimental work on lie detection, with word-association and reaction-time for the next forty years (Binswanger, 1908; Crosland, 1931; Hathaway, 1929; Hunt & Landis, 1935; Jones & Wechsler, 1928; Lockhart, 1975). Until the work of Darrow (1931), researchers only looked at the amplitude of electrodermal responses. The value of the recovery curve has been an available aspect of analysis of electrodermal recordings since that time.

The first substantial use of the galvanometer for law enforcement cases was by Father Walter G. Summers of Fordham University in New York. In several articles (1936, 1937, 1938, 1939), he reported that he had used the Fordham Pathometer (a skin resistance recorder with a range of 2,000 to 150,000 ohms) in over ninety criminal cases and that the results had all been successful, confirmed by confessions, or supplementary evidence. He added that his decisions had established the innocence of persons accused of crimes (1938). In a research project with 271 individuals in which he tested for guilty, accomplice, or innocent status, he was accurate in more than 98% (1939).

An interesting feature of Father Summers' technique is that it was a control question test in which the three relevant questions followed a sequence of events: "Do you know who took the money?", "Did you take the money?" and "Do you have the money on These "significant" questions were asked in that your person?" sequence twice during the recording, interspersed with a larger number of "non-significant" questions of a matter-of-fact, as "Are you wearing a black coat?" and "Did you eat breakfast?" There were also interspersed "emotional standards," as "Were you ever arrested?", "Are you living with your wife?" and "Do you own a revolver?" Summers noted that the emotional standards were selected after a careful analysis of the suspect's life history and after the examination of his psychogalvanic reactions to a preliminary series of questions. He said, "When chosen properly, the emotional standards tend to evoke within the individual rather intense psychogalvanic reactions to surprise, anger, shame or anxiety over situations he would ordinarily prefer to conceal. In the examination of suspects an emotional standard precedes each significant question." Here, then, is the pairing of control and relevant questions, plus interspersed irrelevant questions over twenty years before Backster's introduction of the zone comparison However, Summers' test lacks the sacrifice relevant and test. symptomatic question which are a part of the zone. And, unlike Backster, Summers taught only a few others. The nature of Summers' preliminary test is unknown, so we cannot compare it with the Keeler, Reid, and other preliminary tests.

In terms of analysis of his galvanometer charts, Summers states,

"For purposes of interpretation we contrast and compare the reactions to the significant questions with the reactions to the emotional standards. If the deflections to the critical (significant) questions are consistently greater than the deflections to the emotional standards, the individual is consciously trying to deceive the examiner. If, on the other hand, the deflections to the critical questions are not consistently greater than those to the emotional standards, the individual is truthfully expressing his state of mind. This is the essential criterion of interpretation."

Here we have a clear expression of current control question analysis. We do not know if Summers compared the reactions to relevant questions with the reactions to just the preceding control, or to both the preceding and following control (which preceded the next relevant). In an explanation of his analysis of a chart he simply said that the reactions to questions K (know who), G (did you), and P (money on your person) were "of much greater magnitude than the deflections to other questions within the immediate vicinity." He added, "The individual was correctly adjudged guilty of taking and keeping the money." In a footnote, Summers notes "The height of a deflection is measured from a point where the curve first begins to rise, to the point where it attains its maximum height." In commenting on a murder case, Summers said, "It is quite important to make allowance for the effects of habituation." and "It is equally important to interpret the records in light of the constitutional type indicated by the record." In the analysis of a chart of a person suspected of being an accomplice in a murder, he reports the repetition of the significant question four times and said the record showed a constant diminution in the deflections, the last falling below the levels of the reactions to non-significant questions. Examined three days later, the gradual diminution occurred again, and this was determined by Summers to be indicative of being not deceptive. The results were confirmed by trial and details discovered by the prosecutor (Summers, 1939). Father Summers' professional practice involved over 6,000 tests, cases conducted over a five year span (Block, 1977).

Larson, writing about police cases in 1922 conducted with a recording polygraph with pneumo and cardio tracings, discussed marked changed in the records, deviations "so definite that they can be differentiated from the rest of the record." There may be "an increase or decrease in frequency, a marked depression or excitation, or a more or less summative effect. In all cases of deception yet encountered the curve differs for that of the controls or the person who does not repress." He added that, "The apprehension of an innocent man accused of a crime does not interfere with the test." Larson also noted that the irregularities involved in deception disappear with confession. In an article by Larson in 1923, he included illustrations from charts of criminal suspects with descriptions of the case. Descriptive terms included the "cardiac curve," a "Change in systolic pressure from 149 to 184 mmHg," "tension becomes markedly increased," an "anticipatory effect" (absent in the innocent), "changes discernible in the heart and respiratory curves," and "marked tension may be seen in the respiratory tracing in the tremulousness of the waves in addition to the inhibitory effect caused by lying." Larson noted that systolic blood pressure recordings were much lower after confession. At the end of his

1923 article, Larson lists reactions he had observed that related to deception, and that they may occur in both the cardiac and respiratory curves, or in one alone, but more often in the cardiac action:

- a. Increase in blood pressure a rise
- b. Decrease in blood pressure
- c. Increase in height
- d. Increase in frequency
- e. Summative effects
- f. Incomplete inhibition
- g. Complete inhibitory effect

h. Irregular fluctuations, especially noted at the base of each cardiac pattern

i. Combination of any of the above effects in the same individual

j. These changes may occur with but little latent period, or they may be accumulative in effect and more generally distributed

In 1932, Larson and Haney published a lengthy paper on personality variables in which the data was drawn from their use of the polygraph in criminal cases, supplemented by clinical and psychometric evaluations of their criminal-suspect subjects. The paper also included results of experimental work. Changes that are independent of the stimulus were put in two lists. The 'Chief Changes Noted in the Cardiac Curve' were:

- 1. Increased rate
- 2. Decreased rate
- 3. Increased amplitude of contraction, height and width
- 4. Decreased amplitude of contraction, height and width

5. Rise of the curve from baseline, the height of the contraction remaining unchanged

6. Rise of the curve from baseline, the contraction increasing in height

7. Rise in the curve from baseline, the contraction decreasing in height

8. Fall from baseline

9. Summative or Tetanic effect, two or more contractions running together before a return to the baseline

- 10. Extra contractions
- 11. Refractory periods or inhibitory pauses
- 12. Disappearance of dicrotic pauses
- 13. More than one notch in the individual contraction
- 14. Any other change in the foregoing

15. Quantitative estimation of systolic, diastolic, and pulse taken intermittently or by some continuous method

Larson and Haney next listed, the 'More Important Variations Noted in Respiratory Tracings:'

1. Increased rate in smooth curve or contraction wave

- 2. Decreased rate in smooth curve or contraction wave
- 3. Increased height and width of plain curve
- 4. Decreased height and width of plain curve
- 5. Increased rate and increased height of plain curve
- 6. Increased rate and decreased height on plain curve
- 7. Decreased rate and increased height of plain curve
- 8. Decreased rate and decreased height of plain curve

9. Composite plain curve or any mixture of elements of the plain curve

10. Increased rate of notched, serrated, or tremulous curve

11. Decreased rate of notched, serrated, or tremulous curve

12. Increased height and width of notched, serrated, or tremulous curve

13. Decreased height and width of notched, serrated, or tremulous curve

14. Decreased rate and increased height of notched, serrated, or tremulous curve

15. Increased rate and decreased height of notched, serrated, or tremulous curve

16. Decreased rate and increased height of notched, serrated, or tremulous curve

17. Decreased rate and decreased height of notched, serrated, or tremulous curve

18. Composite plain curve or any mixture of elements of the notched, serrated, or tremulous curve

- 19. Composite, plain, an serrated curve
- 20. Inhibition or an arrest of plain curve
- 21. Inhibition or an arrest of notched curve

In 1936, Leonarde Keeler wrote a chapter, 'The Detection of Deception' in a book on Scientific Criminal Investigation, published by the Northwestern University School of Law. In regard to detecting deception, he writes about peak of tension testing as well as the specific response test (relevant and irrelevant question), and describes how these tests are administered. He said that in criminal cases (as opposed to laboratory experiments) the awareness of the guilty person intensifies the fear of detection, accentuating the bodily The innocent subject will give no symptoms of fear, or changes. if fear exists at the onset of the examination, they will disappear as the examination proceeds. He noted that guilty persons often made an effort to prevent exposure by suppressing changes, which affected "semiautomatic voluntary muscular movements," particularly those of respiration. The innocent, he said, make no effort to control symptoms of fear. Although Benussi mentions countermeasures in his research report (1914), here we have Keeler mentioning their appearance in real cases and the lack of them among the innocent. The appearance of deliberate countermeasures has been suggested as an indication of deception (Magiera, 1975). In 1931 Keeler describes the blood pressure reactions from 4 to 10 mm. of mercury. Usually the diastolic pressure increases considerably more than the systolic. During a test lasting ten minutes, the general pressure may rise from 8 to 20 mm. of mercury, depending on the individual. The respiration of the guilty person becomes more rapid and, in most cases, he attempts to control his responses at periods immediately following deception, tending to shorten both inspiration and expiration and breathe more slowly. In consequence there is an oxygen debit and, on the following question; the breathing becomes deeper and more rapid than the preceding normal. The innocent suspect is not prompted to control his responses, and his respiratory curve becomes more regular as the test These descriptions, and similar material in later progresses. articles were a beginning in the effort to characterize response, which was necessary before specific values could be assigned to responses.

In that 1936 article Keeler had become more specific in his descriptions. In a chapter on 'Detection of Deception' he not only gave the general descriptions cited above about fear, and about countermeasures, he also listed specific criteria for reading a peak of tension test:

Blood Pressure Pulse

1. Peak of Tension - highest point of blood pressure curve

2. Decrease in pulse frequency usually followed by a slight increase

3. Greatest variation in blood pressure curve, immediately following stimulus

4. General irregularity of blood pressure curve preceding point of deception followed by a smoother curve

5. General gradual rise in blood pressure curve following point of deception (rare type of response)

<u>Respiration</u>

6. Regular normal respiration to point of deception, suppression (decreased amplitude and rate) during period between deception stimulus and next stimulus followed by relief (deeper and more rapid respiration)

7. Suppressed respiration during entire period preceding deception stimulus followed by deeper respiration for remainder of test

Here we have the beginning of lists of features that later became lesson plans on chart interpretation. However, Keeler did not start his formal school of instruction until 1948. During the interim, other material appeared on chart interpretation. In books there was William M. Marston's <u>The Lie Detector Test</u> (1938), Fred E. Inbau's <u>Lie Detection and Criminal Interrogation</u> (1942), and Clarence D. Lee's Instruction Manual for the Berkeley Psychograph (1943), which with the addition of some photographs became <u>The Instrumental Detection of Deception: The Lie Test</u> (1953).

William M. Marston, in his book <u>The Lie Detector Test</u> (1938) said that changes in blood pressure are the chief and only dependable criterion of deception. While he did not trust respiration alone, he did observe that when marked changes in respiration tracings accompany change in the blood pressure level, a judgment of deception may safely be made. Two illustrations of continuous records were included to illustrate his points.

In describing a cardio reaction he observed, "The shifting of the entire mass of pulse tracings toward the upper edge of the recording strip." He added that, "Variations in the pulse record may serve to call attention to b.p. level changes but are not in themselves significant, since a compensatory mechanism may decrease the b.p. as the pulse quickens, or again this effect may be absent, equally without relevance to the deceptive reaction."

As for respiration, he observed that "an increase in amplitude of breathing is the most easily noticeable change likely to accompany emotions ... and this increased amplitude becomes suggestive of lying." After mentioning the problem of applying Benussi's ratios and the large amount of work necessary, Marston observed that a sudden "hump" in the breathing record may be meaningful, as may a "shoulder" in either the inspiration or expiration tracing, a sudden irregularity indicating a "catching of the breath," or an "unaccountable flattening out of the whole respiration tracing indicating an extended series of shallow breaths." If Marston had little faith in the significance of respiratory reactions alone, his descriptions suggest he took time to study it. Unfortunately the "hump," "shoulder," and "catching of the breath" are vague.

The first report of real cases in which numbers were assigned to values of reactions is probably that of Professor John E. Winter who reported in 1936 on the results of cardio-pneumo psychogram tests of 25 suspects. His work was confirmed by the post-test confessions of the thief. Professor Winter also administered Jung's word association/reaction time test to the suspects, but it did not perform well. One of 25 women who resided in dormitories at West Virginia University was suspected of stealing from dormitory rooms. The cardio-pneumo psychogram test format involved a "normal or practice curve" as a basis for comparison in evaluating the subsequent three real tests. The scoring system devised by Winter applied to both the cardio and pneumo patterns. He used "0" for no significance, "nothing to indicate guilt." He used "1" for "some significance and points in direction of guilt," and "2" for "distinct signs of guilt." The result was that one woman scored cardio reactions of "2" and "1" in her first two real tests, and confessed.

Winters' innocent subjects were not without reactions, but none reached a level "2" reaction. One woman showed a level "1" reaction in the pneumo pattern of her first two of three real examinations, and two women showed level "1" reactions in the cardio pattern on their first two of three real examinations. The pneumo cleared 24 of 25 suspects, but missed the thief. The cardio cleared 22 of the 24 innocent and identified the thief. Winter reported his findings in the <u>Journal of Applied Psychology</u>, a widely circulated psychology quarterly. However, this appears to be Winter's only work involving deception testing.

Fred E. Inbau, whose experience came from the Northwestern University Crime Lab and the Chicago Police Department Laboratory, wrote <u>Lie Detection and Criminal Interrogation</u> (1942) in which he listed separate criteria for the relevant/irrelevant question test and the peak of tension test. For the RI test he said the "most reliable and definite indication of deception" is a "simultaneous

274

occurrence of a suppression in respiration and an increase in blood pressure immediately after the subject answers a question asked by the examiner." The following list is extracted from the text (pp. 12-16) which was illustrated by chart segments.

Blood Pressure

Increase in blood pressure Decrease in blood pressure Slowing of pulse beat <u>Respiration</u> Suppression Heavier breathing - relief

The responses must constitute a deviation from the subject's normal pattern and the response must be duplicated on at least two different test records.

For a peak of tension test, Inbau's criteria was "the highest point in the blood pressure-pulse tracing," or "a line of demarcation, so to speak, between a somewhat irregular, unsteady portion of the respiration or blood pressure-pulse tracing, and a more regular, steady recording from that point on." Inbau added, "In many instances the "peak of tension" is "attributable to the guilty person's anticipation of being asked the one question on the list to which he will lie," and to "the relief of tension he experiences after answering that question."

In regard to the psychogalvanograph unit, Inbau devoted several pages of text and illustrations. He wrote, "A deflection in the electrodermal tracing ... has been found to be a very accurate and dependable indication of deception in experimental cases." He added that while electrodermal diagnosis was approximately 95% accurate in experiments, it had not been found to be of much practical value, with an occasional exception. In his illustrations, the polarity of the unit was the opposite of the current standard, so decreases in skin resistance were represented by a downward direction of the pen. His criteria was: significance may be attached to the last sizable deflection, or to the lowest point reached by a declining curve.

In a letter from C.D. Lee, then in Berkeley, to John Edgar Hoover, Director of the FBI, dated August 26, 1937, Lee described the various test methods then in use. In describing a peak of tension test that had four correct items in a list of twelve items about a robbery at #3, #6, #9, and #12, Lee said, "If the peaks in the blood pressure curve correlate with questions 3, 6, 9, and 12, there could be little doubt that the suspect was the right man." (Ansley & Furgerson, 1987).

When C.D. Lee published his <u>Instruction Manual for the</u> <u>Berkeley Psychograph</u> (1943), he was an established authority on the polygraph. His polygraph instrument, the Berkeley Psychograph, along with the Keeler Polygraph, were the two most commonly used instruments. Lee set forth noteworthy aspects for chart interpretation, plus symbols to indicate the degree of reaction. In evaluating the "recorded blood pressure changes," he listed:

> Distribution of the reactions Degree of the reaction Trend of the gross curve Rate of ascent of the curve Latent period of reaction Duration of reaction

In discussing "Degrees of Reaction," Lee suggested symbols of "x" and "-" (plus and minus) for degrees, using this table:

<u>Symbol</u>	<u>Name of Symbol</u>	Explanation				
	two minus	fall of curve below basic level				
-	one minus	no change				
x	one plus	slight rise				
xx	two plus	pronounced rise				
XXX	three plus	very pronounced rise				

Illustrations of cardio reactions and differences were included in the text. In regard to individual pulse waves he listed the following:

Rhythm, or regularity

Rate, normally 72 per minute

Amplitude, or distance from one end of the oscillation to the other

Position of dicrotic notch

Lee listed criteria for evaluating respiration:

Rhythm, or regularity Rate, normally 18 per minute Magnitude, or volume Inspiration-expiration ratio (abbreviated I/E) Suppression Holding Notched or serrated strokes

There were also illustrations of these change which appeared in C.D. Lee's 1953 book, <u>The Instrumental Detection of Deception</u>. The book, which received much more attention, added pictures of charts to the diagrams he used in 1943, but contained lists that were quite similar. For recorded blood pressure changes he listed:

> Specificity or distribution of the reaction curves Degree of the individual reaction Rate of ascent Latent period of reaction Duration of reaction General trend

Lee used the same table for the "Degrees of Reaction," with the dashes and "x" marks.

For respiration, Lee listed:

Rate

Magnitude or volume

Inspiration-Expiration ratio

Suppression

Notched or serrated strokes

Lee dropped "rhythm, or regularity" and "holding" from his list of criteria, although the illustrations for the respiration in the book was the same as they used in the manual. For analysis of the pulse waves, Lee used the same list, deleting only "rhythm, or regularity." Lee did not list GSR criteria.

In 1943, Harney questioned whether or not the respiratory responses listed by Trovillo (1942) were indicative of deception, or might also be found in non-deceptive situations. Harney found that the mere act of preparing to rearrange pieces of cardboard into a geometrical figure produced reactions of apnea, suppression, or baseline rise in 65% of the 54 non-criminal adult males. Harney suggested that examiners should "watch for questions of a conflicting nature in which there is a delayed decisions to be made since in these cases respiratory responses apparently symptomatic of deception may appear." Here, he is

describing the undesirable effect when we don't word relevant questions properly; but the effect is what we expect from a control question. Harney didn't mention control questions, but suggested as a further check that the examiner compare the pneumo responses to irrelevant questions to the pneumo responses to the relevant questions before deciding the responses are deception criteria.

The description of a cardio pattern in a peak of tension test was given by Marcuse and Bitterman in 1946. They said, "for example, blood pressure is often found to rise gradually to a point first following the significant detail and then to decrease or remain level until termination of the record."

It was logical that as lists of analytic factors grew, there were attempts to make them systematic, and apply numerical values. Winter first used numbers in 1936, then Lee (1943, 1953) used plus and minus signs, and Backster, while at the Keeler Institute, developed a numerical system emphasizing the cardio (Ansley, Starke R. Hathaway and Clinton B. Hanscom (1958) attempted 1951). to use a statistical method to analyze confirmed polygraph charts from real cases. They measured only the cardio rise following each relevant and each irrelevant question. They did not measure the rise following control questions. They averaged the reactions to the relevant questions and the reactions to the irrelevant question, and subtracted the latter average from the former With a range of -.2 to +.8, they chose an arbitrary average. value of +.5 as a cutoff. This correctly differentiated 88% of the 17 innocent sets of charts but only 49% of the 33 quilty sets of charts. While the false negative rate of 51% was not useful, the authors suggested that some form of statistical treatment was needed to replace the arbitrary methods then in use.

In 1961, Backster published a new test format, the zone comparison. Like Summers' earlier test it featured pairs of control questions and relevant questions, plus irrelevant questions. However, Backster added symptomatic questions, a sacrifice relevant question, guilt complex questions, and an optional suffix ("SKY") asking if the subject suspected (S) who did it, knew (K) for sure who did it, or if he (You) did it. In the following year Backster published a scoring guide for his zone comparison test. His definition of control questions and guilt complex questions added to John E. Reid's introduction to the use of control questions and guilt complex questions in a fixed sequence test in 1947 and improved on Summers' pairing of control and relevant questions in 1939. The Backster zone format was recognized by some examiners as superior to the relevant/ irrelevant and Reid control question formats because the control and relevant questions were paired and the test was designed to cover only one issue at a time. However, the single issue limit

278

was also considered a drawback. Backster's features made scoring easier, particularly when the Backster numerical scoring appeared in print in 1962, and in a notepack in 1963. The new scoring system assigned plus and minus numbers to the comparison of control and relevant question; plus for truthful and minus for deceptive. The system had extensive rules for a variety of circumstances, and defined the requirements for each evaluation. It gave cutoff points for truth and deception scores, with the scores in between being inconclusive. The scoring concept was novel and brilliant.

Although the Army CID adopted Backster's zone comparison format and Backster's scoring system, they changed the scoring rules, cutoff points, and later, the question sequences. Backster has also made changes in his rules and scoring procedures (1979) as well as cutoff scores (1985), but his initial concept is the basic idea behind the scoring of most control question techniques. Among the many control question tests there are several zone comparison tests and three that are widely used: Backster's, Army (now DoDPI), and Utah; the last being the creation of David C. Raskin who received his initial polygraph training at the Backster School of Lie Detection. A comparison of these three scoring methods has been published by Richard S. Weaver (1980, 1985), and there have been a number of other papers examining possible variables in scoring zone charts (Capps, 1991; vanHerk, 1990; Haney, 1972). Also, there is a completely different method for reading zone charts, called Rank Order Scoring or Horizontal Scoring, introduced by Gordon and Cochetti (1987) and Honts and Driscoll (1988). The latter, employing the Timm (1982) line-length for respiration evaluation was designed for computer analysis. Although another method for computer analysis of respiration patterns has been designed by Nakayama and Yamamura (1990), it has yet to be adopted in American computer analysis systems. In addition to Backster, Army, and Utah zones, there is the Canadian Police College version (Kopang, 1985) and the Matte quadri-zone (Matte, 1980). The Canadian method is widely used in Canada and is also used by some U.S. law enforcement agencies because the Canadian school has generously trained examiners from the United States. The Matte method, the topic of one book (Matte, 1980) and some research (Matte & Reuss, 1989, 1992), is not widely used. All of the systems have adopted the Backster seven-point scale, but with different rules and cutoff scores.

<u>Comparative</u> Studies

There have been several studies comparing various ways of scoring charts. Some compared numerical scoring systems. For example, Weaver (1985) compared the effects of scoring fifteen sets of control question charts by the three most common methods: Army (USAMPS), Backster, and Utah. While USAMPS (U.S. Army Military Police School) and Utah agreed in 14 of 15 decisions, Backster and USAMPS agreed in only ten cases. The Backster system produced nine inconclusive opinions compared to four for USAMPS and five for Utah. A study by vanHerk (1990) compared a three-point system with the more conventional seven-point system and found the three-point system produced more inconclusive However, when he lowered the cutoff scores from +/-6 to results. +/-4, the inconclusive rate was reduced. Capps and Ansley (1992) had similar results with the three-point system, a high inconclusive rate with +/-6 but a favorable rate at +/-3. Using a modified zone comparison test for 56 mock screening examinations, Barland (1981) evaluated the charts by the USAMPS method of +/-3or each question, where the relevant question was compared to the nearest control, scored again, but with the five relevant questions scored against the strongest control response on the chart, and scored a third time by the relevant/irrelevant method of looking at the size and consistency of responses without reference to the controls. The RI, or global, method was correct on 86% of the guilty and 75% of the innocent; the USAMPS method was correct on 81% of the guilty and 76% of the innocent; and the strongest control method was correct on 68% of the guilty and 83% of the innocent.

Thomas (1980) reported on the analysis of eight polygraph examinations (involving a credit union theft of \$2,000) in which all of the examinations were worded and conducted much alike. The charts were numerically scored by the original examiner and another examiner, Richard Weaver of the Wisconsin Crime Laboratory. The two examiners were graduates of different schools and had learned different methods of numerical evaluation, and used different cutoff scores. Nonetheless, both arrived at conclusions of truthful for the same six of the eight subjects, and inconclusive for two of the subjects.

Crowe, Chimarys and Schwartz (1988) used thirty sets of mock crime polygraph charts conducted in the General Question Test format, a control question technique. GQT is not a widely used technique. The thirty sets of charts were independently scored by nine polygraph examiners. Seventeen subjects were programmed deceptive and 13 nondeceptive. The charts were scored by three 1) comparison of the strongest control reaction to the systems: reaction for each relevant question, 2) comparison of the weakest control reaction to the reaction for each relevant question, and 3) an overall visual or global evaluation of the charts. Discounting inconclusive results, the strongest control was correct in six of seven deceptives (86%) and seven of seven nondeceptives (100%), for a total of 13 of 15 (93%). The strongest control resulted in an inconclusive rate of 16 of 30 (53%). The weakest control was correct in 17 of 17 deceptives (100%), but was correct for none of the nine nondeceptives (0%),

for a total of 17 of 26 (65%). The weak control resulted in four inconclusive decisions (13%). The overall or global approach was correct in eight of ten deceptives (80%), and three of five nondeceptives (60%), for a total of 11 of 15 (73%). The inconclusive rate was 12 of 30 (40%). In this system, the method employing the strongest control was the most accurate, but had a high inconclusive rate, over half. The weakest control was unsatisfactory as it missed all the nondeceptive subjects. The global method was second to the strong control, but it also suffered from a high inconclusive rate.

In order to see how the lists of reactions developed, we have described the teaching of chart interpretation at various The Keeler Institute was the first formal course of periods. instruction, so the teaching of chart interpretation is described there in 1949 and 1951, and the Keeler technique as modified by Leonard Harrelson is described by Harrelson in 1964 and by Ferguson in 1966. The list of reactions taught at the Army course in 1954 was described by Captain Joseph in 1957, and described again in a DoD Polygraph summary sheet in 1991. There is little Many of the other polygraph courses have copied the difference. Army or DoDPI list, and an example of that is given in the description of the Texas A&M course of 1984. Even the Japanese lists and descriptions are similar to ours. It is the Army-DoDPI list, so long in use, that we chose to us as a beginning from which we developed the list used in our research. This review of the courses and their teaching should provide some historical insight.

The Keeler Institute - 1948 to 1951

From 1948 to 1951 the only organized course of instruction on polygraph technique was the six-week course at the Keeler Institute in Chicago, Illinois.

A set of 39 pages of neatly typed notes exist from the April 18 to May 9, 1949 class (Clinchard, 1949), but the notes on chart interpretation are brief. Polygraph instructors in that course were Leonarde Keeler, Jack Harrison and LTC Ralph Pierce, USA Ret.

In regard to pneumograph, Pierce observed that "Pleasure makes the respiration weaker and faster, displeasure makes the respiration stronger and slower, excitement makes respiration stronger and faster, and acquiescence makes respiration weaker and slower." We note that this is directly from the 1908 book by Hugo Munsterberg, <u>On the Witness Stand</u>. Pierce may have mentioned the source but it isn't in Clinchard's notes. Pierce also said any changed from the normal could be indicative of deception, and persons may "increase, decrease, raise, or lower [on the] chart pattern, depending on the individual.

Jack Harrison lectured on the cardiograph and observed that, "Some subjects' pressure rises with pertinent questions; on others it decreases; both could be indicative of guilt. A fast pulse, if continued, could be indicative of guilt or of nervous tension." There are no other notes on chart interpretation.

The 1949 course had Ray Holmes of Associated Research teach instrumentation, two professors taught psychology, one taught physiology, and Dr. Lemoyne Snyder taught homicide investigation.

For the class of February 26 to April 6, 1951 (Ansley, 1951), there were three primary instructors in technique, plus supplementary instruction from an attorney, a medical doctor, a psychologist, a physiologist, a representative of the instrument manufacturers (Associated Research and Stoelting Company), and practicing examiners associated with the Institute. The six-week course became a model for almost all courses that have followed, and was closely copied by the Army CID when they established their course on lie detection at Camp Gordon, Georgia, later in 1951.

The three primary instructors for the February to April 1951 course were Jack Harrison (President of Leonarde Keeler, Inc. - a former Army CID agent), Cleve Backster (who had come from Washington, DC, where he had established the CIA polygraph program), and Albert Breitzman, an examiner in the Evanston (IL) Keeler had died in September 1949, and Jack Police Department. Harrison became President of the company. The instructors used lesson plans, and the course followed a well established pattern. The time devoted to chart interpretation involved lectures and practical work with charts from real cases and with charts students conducted in mock crime cases. All three instructors were involved in teaching chart interpretation. The primary instruments were Keeler Model 302, which had an electrodermal, respiratory and cardiovascular channel.

In regard to the electrodermal, it was taught that:

Large change with a lie

Sensitivity diminishes as tests are continued

After a reaction it may not return to the prior baseline

Reaction time is between one and five seconds

If the pattern reverses from a down-drift it is a strong reaction

Leveling out from a down-drift is a reaction No known inverted (downward) reactions

The pneumograph reactions were described as:

Changes in amplitude Baseline changes Suppressions (at exhalation) Blocking (at inhalation) Disturbed pattern (irregular breathing) Several deep breaths at point of deception

Break in the pattern The cardiosphygmograph responses were given the most emphasis, and described as:

All variations are possible reactions Slowing pulse rate or an increase in pulse rate Rise and fall of the systolic and diastolic Extremely rapid pulse rate (110 bpm or faster) A continuous rise after stimulus

A normal pattern followed by up to three 'bounces' (rise and fall)

Also noted as unusual reactions are changes in the dicrotic notch (alone), an increase in the stroke size followed by a decrease, a drop in the pattern without a preceding rise, extra systoles only on the relevant questions, irregular pulse rate up to the stimulus followed by a regular pulse, an irregular pulse rate after the stimulus, an increase in pulse rate after the stimulus.

What is particularly interesting about the 1951 Keeler instruction is the introduction of a numerical system by Cleve Backster. Responses in the cardio channel were graded 0, 1, 2, 3, or 4, with 4 indicating the subject's largest response, and the other number grades in proportion. A plus or minus could be assigned for finer distinctions. As for the pneumograph, it was to be used to influence the value given to cardio reactions and a strong block or suppression could add a point; so a two-point cardio reaction (half of the largest cardio reaction) could be three points if accompanied by a strong pneumo response. As a general guide Backster recommended a weight of 75% cardio, 20% pneumo, and 5% electrodermal.

The Army Course - 1954

In 1957 Professor V.A. Leonard published the first of two volumes of papers on polygraph testing by various experts. Captain C.N. Joseph, U.S. Army, was among the eleven contributors

283

to the first volume. Captain Joseph, noted the editor, had for a number of years been associated with the Provost Marshal General's School at Camp Gordon, assigned to the Lie Detector Training Section. His paper, "Analysis of Compensatory Responses and Irregularities in Polygraph Chart Interpretation" was originally presented at the 1954 seminar of the American Academy of Polygraph Examiners in Washington, DC. In his paper, Captain Joseph listed the following irregularities in the polygraph recordings as generally accepted as being indicative of deception if they appear only at relevant and/or control questions.

In the pneumograph the following basic changes may be indicative of deception:

- a. Changes in rhythm or regularity
- b. Changes in amplitude or volume
- c. Changes in inhalation-exhalation ratio
- d. Notched or serrated inhalation or exhalation stroke
- e. Changes in baseline
- f. Loss of baseline
- g. Hyperventilation
- h. Suppression
- i. Holding or blocking
- j. General pattern changes

In the galvanograph tracing the following basic changes may be indicative of deception:

- a. Vertical rise at the point of deception
- b. Double saddle response

c. Long duration of responses following point of deception

d. General pattern changes

e. Drops at point of deception (in non-centering galvanometer)

In the cardio-sphygmograph tracing the following basic changes may be indicative of deception:

- a. Increase in blood pressure
- b. Decrease in blood pressure
- c. Increase in pulse rate

- d. Decrease in pulse rate
- e. Increase in amplitude
- f. Decrease in amplitude

g. Change in position or disappearance of the dicrotic

notch

- h. Extra systoles
- i. Distribution of the reactions
- j. Degree of the reactions
- k. Trend of the gross curve
- 1. Rate of ascent of the curve
- m. Latent period of reaction
- n. Duration of reaction

o. General pattern changes <u>Keeler Institute - 1964</u>

In his book, <u>The Keeler Technique</u> (1964), Leonard H. Harrelson outlined rules for chart interpretation. The book was the text at the Keeler Institute from 1964 to 1990, and the outline may well have been used before 1964.

For the pneumograph pattern he listed the following 'reaction types':

- a. Suppression
- b. Apnea (at end of inhalation or exhalation)
- c. Change in baseline only, downward
- d. Change in top of tracing only, upward
- e. Change in rate
- f. Change in I/E ratio
- g. Labored breathing
- h. Sustained rapid respiration
- i. Involuntary movement
- j. Momentary hesitation of the pneumo pen (pen stop)
- k. Erratic tracing throughout
- 1. Scaling tracing

285

For the galvo pattern, 'types of reactions':

a. Single rise and mechanical return

b. Single rise and normal return

c. Rise, mechanical return, second rise and mechanical return

For the cardio-sphygmograph pattern, 'reaction possibilities':

- a. Change in position of dicrotic notch
- b. Change in amplitude
- c. Change in rate
- d. Single rise and return (single bounce)
 - 1) Rise and return in systolic only
 - 2) Rise and return in diastolic only
 - 3) Rise and return in both systolic and diastolic

e. Double rise and return in the cardio tracing (double bounce)

1) Rise and return with secondary rise and return encompassing any or all of the three pressure components

2) Rise and partial return with secondary rise and return encompassing any or all of the three pressure components

f. Triple rise and return in the cardio tracing (triple bounce)

- g. Sustained rise
- h. Skip beat
- i. Sustained rapid pulsation

1) a pulse rate of 100 or more may indicate deception in and of itself

- j. Inverted
 - 1) This is rare, but can occur
- k. Momentary hesitation of the cardio pen (pen stop)
- 1. Involuntary movements by the subject

<u>Relevant/Irrelevant Scoring - 1966</u>

In a book, <u>The Polygraph in Private Industry</u> (1966), the author listed reaction types with illustrations, in Chapter 8, "Chart Interpretation; Reaction Types." The author, Robert J. Ferguson, Jr., was a graduate of the Keeler Institute and was describing the analysis of relevant/irrelevant technique charts.

For the pneumograph he listed the following 'reaction types':

- a. Suppression of breath, top line
- b. Blocking (apnea), top line
- c. Suppression of breath, baseline
- d. Blocking (apnea), baseline
- e. Erratic tracing throughout
- f. Scaling tracing
- g. Sustained rapid respiration
- h. Continuous labored breathing (hits pen stops)
- i. Changes in baseline, only downward
- j. Changes in top line, only upward
- k. The three-stair step
- 1. Changes in rate per minute
- m. Suppression of both top and baselines

n. Subject movement affecting pneumo pattern For the galvo he listed the following `reaction types':

- a. Single rise and fall
- b. Rise and fall of long duration
- c. Sustained rise
- d. Upward drift of long duration
- e. Rise and secondary wave
- f. Downward drift
- g. Abnormal galvo tracing indicating subject movement
- h. Double saddle reaction

i. Reaction to all questions up to, and including, a crucial question with little or no response thereafter

j. Little or no reaction up to the crucial question with more overall response thereafter

k. Upward drift, or rising baseline up to the crucial question, and relief with downward drift thereafter

1. Reaction with rising baseline up to a crucial question, and a sustained reaction thereafter

m. Reaction to all questions with the sharpest or greatest reaction occurring at a crucial question

n. Reaction to all questions with the longest duration occurring at, during, or following a crucial question

For the cardiosphygmograph he listed the following `reaction types':

a. Single rise and return in systolic, diastolic, or both

b. Double rise and return, systolic, diastolic, or both

c. Rise and partial return with secondary rise and

return

- d. Triple rise and return, both systolic and diastolic
- e. Sustained rise, systolic and diastolic
- f. Skip beat (extra systole)
- g. Sustained rapid pulsation
- h. Inverted reaction, systolic or diastolic (rare)
- i. Involuntary subject movement

Texas A&M Course - 1984

The following is from an outline given to students at the Texas A&M University Polygraph Examiners' School class of August 6 to September 28, 1984. It was entitled "Summary: Terminology and Interpretation of Test Graphs," and had 74 typed pages, with most of that devoted to segments of charts illustrating each kind of response. We list this as an example of what many schools have done, copy the Army or someone else's list. The outline stated:

Pneumograph changes which may be indicative of deception:

- a. Change in rhythm or regularity
- b. Change in amplitude or volume
- c. Change in inhalation-exhalation ratio
- d. Notched or serrated inhalation or exhalation stroke
- e. Change in baseline
- f. Loss of baseline
- g. Hyperventilation
- h. Suppression
- i. Holding, blocking, or apnea

Galvanic tracings which may be indicative of deception:

- a. Vertical rise at point of deception
- b. Double saddle response

c. Long duration of response following point of deception

Cardio tracings which may be indicative of deception:

- a. Increase in blood pressure
- b. Decrease in blood pressure
- c. Increase in pulse rate
- d. Decrease in pulse rate
- e. Increase in amplitude
- f. Decrease in amplitude
- g. Change in position or disappearance of the dicrotic

notch

h. Extra systoles

Other factors in evaluating blood pressure tracings:

- a. Distribution of the reactions
- b. Degree of the reactions
- c. Trend of the gross curve
- d. Rate of ascent of the curve
- e. Latent period of reaction
- f. Duration of reaction

Inasmuch as Texas A&M taught several techniques, it reproduced exactly the rules for Backster Zone Comparison when teaching that test format. The rules and illustrations were from the Backster School of Lie Detection.

<u>Japanese Literature</u>

The concept that reactions can be classified into groups has spread to all polygraph operations. The Japanese, who read the English-language publications on polygraph testing, have published a number of articles on deception criteria. For example, Kizaki and Osako (1979) writing in Japanese with English titles under illustrations of respiratory reactions listed: rate changes, baseline changes, staircase, suppression, and block. Of eight references in their article, four are to American publications:

Arther, 1971; Cutrow et.al., 1972; Reid and Inbau, 1966; and Thackray and Orne, 1968. Similarly, the Japanese have published other material on the classification of reactions in each of the components:

Cardio: Kasai, 1965; Kawakami, 1968; Suzuki et al., 1977; Suzuki and Watanabe, 1972; Wakamatsu and Yoshizumi, 1968; Watanabe and Kijitani, 1973; Watanabe and Suzuki, 1977; Watanabe et al., 1981; Yamamura and Sasaki, 1990; and Yamaoka and Suzuki, 1973.

Electrodermal: Kosugi and Yamashita, 1978; Miyaki, 1978; and Suzuki et al., 1973, 1974.

Respiration: Nakayama and Yamamura, 1990; and Inagaki and Kira, 1978.

Japanese articles have also discussed classification of reaction patterns for all channels: Adachi and Suzuki, 1990; Imamura et al., 1960; Miyake, 1978; Ohnishi et al., 1976; Shimizu, Hayashi and Yamaoka, 1978; Suzuki and Hikita, 1964; Wada and Yamaoka, 1976; and Yamaoka and Suzuki, 1973.

While the Japanese lists of reaction patterns, and their descriptions, are partly a product of the Western polygraph literature, and partly based on their research and empirical observations, they have had no perceptible influence on the discussion of deception criteria in the U.S., Canada, and Israel, It is unfortunate that Japanese polygraph research is not more widely available, as the quality is excellent.

Backster Zone Comparison

The <u>Backster Zone Comparison Technique: Chart Interpretation</u> (Summary) was distributed by Cleve Backster at the 20th annual seminar of the American Polygraph Association in Reno, Nevada, on August 7, 1985. He listed as 'Primary Reaction Indications,' the following:

Breathing

- a. Suppression
- b. Baseline arousal
- c. Apnea
- d. Slowing in exhalation rate

GSR

a. Arousal

Cardio

- a. Blood pressure arousal
- b. Amplitude reduction
- c. Rate reduction

Backster added a section on "primary relief indicators": Breathing - hyperventilation; GSR (no primary indicator), and Cardio - amplitude increase and rate increase. The summary was accompanied with illustrations, the sequence of questions for his own test, Reid, Arther, Army, and Keeler-Harrelson, and rules for applying numerical scoring. His initial system of dropping the least productive channel score (Backster, 1963) was later changed (1979), and Backster now scores all three channels. The cutoff scores were changed in 1984, and the rules have remained rather constant since then (Backster, 1990).

The Backster rules are written so that only those trained by Backster, or those familiar with his terminology can understand them. We have translated these into the more common polygraph terms: These are taken from his 1990 rules.

1. To arrive at a score of +2 or -2, there must be a significant and timely tracing reaction in either the control or relevant question zone. If there is no reaction to the relevant question, it should be compared with the neighboring control question reaction that has the larger timely reaction. If there is a timely and significant reaction to the relevant question, it should be compared to the reaction to the neighboring control question containing no reaction or the weakest reaction.

2. Each of the three tracings should be independently assigned spot analysis numerical values. Reactions in other tracings will not be allowed to influence the value in the tracing being considered, but distortions in these other channels may be noted.

3. If a "yes" answer is given to a control question, when a "no" was expected, that control question cannot be used in the spot analysis. However, it can be used if there is a lack of reaction or a smaller reaction to the relevant question is indicated. Also, any comparison with the "yes" answered control response is to be avoided if there is another adjacent and properly answered control question to which there is a lack of reaction. This would not preclude the evaluation of a "yes" answer control if the "yes" was given at the time of the question review during the pretest interview, i.e., approved in advance.

4. No numerical spot analysis can be made if an amplifier sensitivity was changed during the relevant question zone or the control question zone.

5. A reaction is timely if it starts following the first word of the question being asked, or if it starts prior to five seconds past the answer.

6. Regardless of its intensity, any reaction to a relevant or control question which starts prior to the first word of the question that follows can be assigned no more than a +1 or -1 score. This is called an anticipatory reaction rule.

7. If the tracing following a control or relevant question exhibits a presence of relief, or the relief starts prior to five seconds after the answer, by deduction that relief can be classified as a lack of reaction. [Relief indications are hyperventilation, a cardio amplitude increase, and a cardio rate increase. There is no relief pattern for the electrodermal.]

8. If a cardio reaction has not completed its recovery within a relevant question zone or a control question zone, and continues into the zone of the question that follows, by deduction that recovery response may be considered a lack of reaction to that following question. However, no more than a +2 or -2 may be assigned if the lack of reaction to the following question is part of a numerical spot analysis.

9. If the GSR pattern doesn't show any reaction, then a small measured reaction is assigned for the purpose of comparison. The measured unit assigned to the GSR depends on the chart scale and is specified in the rule one unit for one-eighth inch, two units if the scale is one-tenth inch, and four millimeters if the scale is metric.

10. If a GSR pattern is plunging in response to either the control question, the relevant question, or both, no numerical value is to exceed +1 or -1. A downward angle of 45 degrees is plunging.

11. If the reaction to a control question is at least four times greater than a minor reaction to the relevant question, this pair should be used for spot analysis rather than a pairing of the weak relevant question reaction with a neighboring control question reaction that is less or shows no reaction.

12. When no significant reaction occurred to a control or relevant question pair, a spot score of no more than +1 or -1 may be assigned based on the overall trend of the tracing average. This is principally done with the cardio pattern but may be done with the respiratory or electrodermal patterns.

13. When a relief pattern begins prior to five second past the point of answer, and where there was no reaction to the preceding question, by deduction it can be inferred that some unrecognized reaction did occur to that prior question. If the relief reaction starts more than five seconds after the point of answer, and there is no reaction to the question being answered, it can be inferred that some unrecognized reaction occurred. Numerical values of no more than +1 or -1 may be used for deduced reactions, and deduced reactions occur only in the cardio and pneumo patterns.

14. When a single breathing cycle does not conform with three other cycles to meet the requirements for a four-cycle horizontal trend, the nonconforming cycle can be counted as conforming if it is not distorted. This rule may not be used to evaluate baseline arousal patterns and may not be assigned a value of more than +2 or -2.

15. Before upgrading a +2 or -2 to a +3 or -3, the two zones being compared must have a minimum of 20 seconds and a maximum of 35 seconds each. That is measured from the first word of one question to the first word of the next question.

16. You cannot upgrade a score of +2 to +3 or -2 to -3 if there is a distortion in the response to either (or both) the relevant or control question. Recentering a pen is, for this purpose, considered a distortion.

17. To upgrade a +2 or -2 to a +3 or -3, the intensity of the reaction must be significantly more dramatic than that required in the "either/or" rule. For the GSR, that +3 or -3 requires a ratio of in excess of four-to-one.

18. If a subtle one or two cycle suppression occurs while either the control question or the relevant question is being asked, this constitutes a mild listening reaction. If this occurs during the asking of both questions it is a mild listening distortion. Lacking other variations in the pneumo pattern following the question, the mild listening reaction does not warrant a +1 or -1 score.

19. If in answering either a control or relevant question there is vocal emphasis that caused a loss in residual air volume, this is a mild reaction. If it occurs in answering both questions, it is a mild distortion. Lacking other variations in the pneumo pattern, the mild answering reaction does not warrant a +1 or -1 score.

293

20. Where there is a cyclic cardio tracing throughout most of the chart, stabilization following a control or relevant question may be scored as +1 or -1. Be sure the stabilization did not result from suppressed breathing, which in turn stopped a vagus pattern.

21. A cluster of extra systoles occurring primarily within the same zone is more likely relief than reaction. A single extra systole followed by a blood pressure arousal should be treated as a distortion rather than a reaction.

DoD Polygraph Institute - 1991

In 1991, the U.S. Department of Defense Polygraph Institute listed the following on their Summary Sheet (IN 330) on Chart Interpretation:

Pneumograph changes from the individual's norm which may be considered indicative of deception are:

- a. Changes in rhythm or regularity
- b. Changes in amplitude or volume
- c. Changes in inhalation/exhalation strokes
- d. Notched or serrated inhalation or exhalation strokes
- e. Changes in baseline
- f. Loss of baseline
- g. Hyperventilation
- h. Suppression
- i. Holding or blocking

Galvanic tracings which be indicative of deception are:

a. Vertical rise at point of deception

b. Double saddle response

c. Long duration and/or degree of response following point of deception

d. Plunging galvanograph at point of deception

Cardio tracings taking the forms of specific responses which may be considered indicative of deception:

- a. Increase and decrease in blood pressure
- b. Increase only in blood pressure

c. Decrease only in blood pressure

d. Increase in pulse rate

e. Decrease in pulse rate

f. Increase in amplitude

g. Decrease in amplitude

h. Change in position or disappearance of the dicrotic

notch

i. Extra systoles

Factors other than specific responses to be considered as possible deception when evaluating charts are:

a. Distribution of reactions

b. Degree of reactions

c. Trend of the gross curve

d. Rate of change of the curve

e. Latent period of reaction

f. Duration of reaction

Interpretation of peak of tension test charts:

a. An increase to point of deception then a level tracing

b. Decrease to point of deception then a level tracing

c. An increase to point of deception and then a decrease

d. Decrease to the point of deception and then an increase

e. Level tracing to the point of deception and then an increase or decrease

f. Erratic to the point of deception and then a smooth tracing

g. Smooth to the point of deception and then an erratic tracing

h. Any change that may occur at the relevant question

The similarity to the 1954 Army list is striking. The more recent list deletes the "General pattern change" at the end of the list for each component, and that is the only significant difference.

295

The Most Productive Channels

Although examiners have generally agreed that the cardio, electrodermal, and respiratory channels are not of equal value in reaching a conclusion of truth or deception, they have not agreed on which is the greatest or the poorest, much less on what value should be given to each one. The current numerical systems are a simple compromise, they give equal weight to each parameter. It didn't start that way. Originally, Backster scored only the two most productive channels, and ignored in scoring the one that was either least reactive or the one in which the reactions were at odds with the other two channels (Backster, 1963). However, he later changed that, and his current rules (Backster, 1991) give equal value.

In preparing for computerization of polygraph data it became obvious that this equal distribution of value needed to be reconsidered. Several studies were undertaken to determine the productivity of each channel, and all used chart sets from control question examinations of specific issue tests. Some were based on field examination results (Barland & Raskin, 1974; Ben-Ishai, 1962; Bowling, 1978; Buckley & Senese, 1991; Elaad, 1985; Elaad et al., 1988; Franz, 1989; Jayne, 1990; Matte & Reuss, 1989; Ohnishi et al., 1976; Rafky & Sussman, 1985; Ryan, 1989; Slowik & Buckley, 1975; Suzuki et al., 1973; Winter, 1936; Yamamura, 1980; and Yamamura & Sasaki, 1990). The data from most of these studies has been summarized by Capps (1991). There have also been a number of laboratory studies that have contributed to the concept of different values from each of the three channels: Berrien and Huntington, 1943; Cutrow et al., 1972; Grimsley and Yankee, 1985; Kubis, 1973; Miyake, 1978; Podlesny, Raskin and Barland, 1976; Stern et al., 1981; Suzuki & Hikita, 1964; Thackray and Orne, 1968; Timm, 1989; Waid, Orne, Cook and Orne, 1981; Yamaoka and Suzuki, 1973; and Yankee and Grimsley, 1986. There is considerable disagreement in the findings about the utility of each of the channels of information. Moreover, none of the studies told us which specific reactions were used the most, and which were the most useful. Without these details our knowledge is incomplete on a vital topic. There are, however, a few studies that tell us something about the frequency of specific reactions.

Frequency Studies

In 1942, Trovillo, at the Chicago Crime Laboratory, studied their confirmed cases to determine where the reactions occurred. Of 23 identified, 11 were blood pressure, six were respiratory, and six were electrodermal.

Carl W. Jensen (1981) took a sample of 66 control question tests and counted the cardio reactions to 502 relevant and control

question. The percentage is the number of times the reactions occurred divided by the number of questions. The list he described as "the standard recognized specific reactions criteria recorded and evaluated from the cardiograph tracing."

	<u>Cardio Reaction</u>	Frequency	<u>Percent</u>
a.	Increase and decrease in blood pressure	363	72%
b.	Decrease in pulse amplitude	326	65%
c.	Increase in blood pressure	172	34%
d.	Increase in pulse amplitude	52	10%
e.	Change in position of dicrotic notch	51	10%
f.	Decrease in blood pressure	48	10%
g.	Increase in pulse rate	43	9%
h.	Decrease and increase in blood pressure	42	8%
i.	Decrease in pulse rate	20	4%

Jenson's list adds to the common lists, "decrease and increase in blood pressure" and deletes "extra systoles." It may be that none of these 66 persons created extra systoles.

Robbins and Penley (1974, 1975) studied 76 deceptive sets of charts (1974) and 140 nondeceptive sets of charts (1975). All of the nondeceptive (NDI) cases were confirmed, and 50 of the 76 (66%) deceptive (DI) were confirmed. In their study of these sets of charts, the authors decided which of the three components were most influential in decision making, the frequency with which different patterns appeared in each channel, and the frequency of apparent attempts to distort the recordings.

As to which channel was most influential, these were their findings:

	DI	<u>(n. 76)</u>	NDI	<u>(n. 140)</u>
Blood pressure - pulse	22	29%	61	44%
Respiration	48	63%	56	40%
Skin residtance	_6_	8%	_23	16%
	76	(100%)	140	(100%)

In their study of the frequency with which specific patterns appeared, the authors only tell us that a pattern appeared at least once in a set of charts. So we know that respiratory suppression followed by relief occurred once or more often in 71 of the 76 sets of DI charts, or in 93% of the sets of charts. We don't know how often that reaction occurred within each set of

charts. (Therefore, the percentages don't add up to 100.) Some tests showed two or more types of reactions in each channel. Here are the findings for respiration:

	DI	(n. 76)	<u>NDI (n</u>	. 140)	
	<u>n. set</u>	<u>s/% sets</u> :	<u>n. sets</u>	<u>/% sets</u>	3
Suppression, followed by relief	71	93%	133	95%	
Apnea	10	13%	6	48	
Change in baseline	not	listed	29	21%	
Change in I/E ratio	5	78	9	6%	

In regard to skin resistance response, they did not list the type of response. Rather they listed the extent from none to excellent:

	<u>DI (n. 76)</u>		<u>NDI (r</u>	<u>n. 140)</u>	<u>140)</u>
	<u>n. set</u> s	<u>s/% sets</u>	<u>n. set</u> s	<u>s/% sets</u>	
Good or excellent GSR response	53	70%	95	68%	
Some GSR response	20	26%	40	28%	
No GSR response/any test chart	_3	4%	5	48	
	76	100%	140	100%	

In writing about cardio responses they gave descriptions of responses and how often they appeared one or more times in a set of charts:

	<u>DI</u> n. set	<u>(n. 76)</u> ts/% sets	<u>NDI (n</u> n. sets	<u>. 140)</u> /% sets
Increase and decrease in blood pressure	48	63%	127	91%
Increase in blood pressure	18	24%	14	10%
Slight change in blood pressure or no change	10	13%		
Pulse rate change	8	12%	6	4%
Change in dicrotic notch	no	details	16	11%

In regard to distortion, there were no attempts by NDI subjects. However 13 of the 76 DI subjects (17%) attempted distortion in one or more relevant charts, and five of the 68 DI subjects (7%) who were administered stim tests attempted to distort the results.

These frequency studies begin to tell us something more about the kinds of reactions that occur in each channel of information
and how often they occur, but they do no tell us when those reactions were used in making spot decisions of truth or deception, or how much numerical weight was given to each reaction. To do that we needed a survey designed for that purpose. We also wanted to know if there is a difference in the way examiners score charts based on their basic training course. No summary has covered such differences.

Rule Variations

In writing history of scoring we have found it impossible to locate and document the numerous rules and changes in rules as they are taught at schools, seminars and meetings. That is not to say they are unimportant. The results of new rules or rule changes have rarely been studied, and the reasons for new rules or rule changes are seldom documented.

Some examples must suffice. In the scoring of Backster Zone Comparison charts the original method (Backster, 1963) required the examiner to total each channel, then remove from the scoring the channel that went away from the trend. However, some years later that rule was deleted, and by 1979 the total scores were always completed (Backster, 1979). Backster's rules compared the relevant question reactions to the weakest adjacent control questions (with exceptions), but when the Army adopted the Backster technique they compared relevant question reactions to the strongest adjacent control. The difference may be significant and is one of the objectives of this research. There are further subsets of these rules; for example, in 1979 the Army school taught rules for all students, with some minor differences for Army CIDC that did not apply to students from other agencies. For example, The Army CIDC rule said you could compare relevant question #7 to control question #6, except when the reaction to symptomatic question #8 is greater than #6 or #7. If it was greater, no evaluation was made of that spot. There were other minor variations for Air Force and Navy (Koll, 1979). Backster created cutoff scores of +/-9 for two charts and higher for more charts. When the Army adopted the zone they switched to a +/-6, regardless of the number of charts. The reasons for the change The difference in numbers of conclusive are not documented. results is probably significant. Later Backster lowered the cutoff scores for the truthful, making the range +5 or -9. One of the objectives of this research was to determine optimum cutoff scores. That has been done by others (Capps & Ansley, 1992; Elaad & Kleiner, 1990; Raskin, Barland & Podlesny, 1978; Raskin & Hare, 1978; Shterzer & Elaad, 1985; vanHerk, 1990), with varying results. It is clear that what is taught and what is practiced in the way of cutoff scores is important. It appears that none of the scores are based on the results of research; rather, they are based on impressions from experience, which is the basis of many of the rules in polygraph testing.

Method

To have an adequate base from which many variables could be considered, twelve examiners were asked to participate in a detailed project of chart review evaluation and criteria identification. We engaged twelve in hope that ten would complete the project. As it was, eleven examiners completed the work, one resigning before reviewing any charts. There were three groups of examiners represented in the selection: federal government, law enforcement, and private practice. The basic polygraph training of the examiners was diverse in that they received training from eight schools: Argenbright International Institute, Backster School of Lie Detection, Gormac Polygraph School, Keeler Polygraph Institute, Munford Institute, Reid College, Southwest School of Lie Detection, and the United States Army Military Police School. All of these basic examiner courses were accredited by the American Polygraph Association, which meant that all had or exceeded a core curriculum of specific topics, and their instructors had degrees and training in their specialties. All of these examiners had also participated in additional training by attendance at one or more American Polygraph Association annual seminars. The participating examiners were white and male, with experience ranging from one year and ten months to 39 years and nine months, when the project started. The mean was 14 years and three months. The experience level of the examiners in terms of specific issue polygraph examinations, based on their estimates, ranged from 80 to 7,500, with a mean of 2,160. All the examiners were either licensed by their respective state licensing boards or, in the case of federal examiners, certified by their agencies. Four of the examiners had some college education but had not received a degree, five had at least a baccalaureate, some with graduate courses, and two had graduate degrees. All had received formal training in one or more of the methods for numerical analysis of control question charts. The study period lasted from February 1991 to February 1992. Each examiner evaluated the same collection of 40 sets of confirmed zone comparison polygraph charts. There were an equal number of male and female subjects. The sets of charts all represented specific issue examinations of which 17 were confirmed as truthful and 23 were confirmed as deceptive. Confirmation of the deceptive cases was by subsequent confession of the examinee, and confirmation of the truthful was by subsequent confession of someone other than the examinee, with investigative details precluding any secondary involvement by the The evaluators were blind to case facts, scores, examinee. conclusions, and all demographic information. They did not know how the questions were worded but control and relevant questions were identified by numbers.

The reason that an unequal number of truthful and deceptive cases was selected was threefold. Actual field polygraph work

does not provide examiners with an equal distribution of truthful and deceptive subjects; examiners who kept track of their decisions might be tempted to be influenced, particularly in their last few cases, to make calls in keeping with an equal distribution; and the third reason had to do with the method of selection of cases. The 40 sets were selected from a file drawer collection of 100 confirmed zone comparison polygraph cases, taking the first 20 male subjects and the first 20 female That provided the unequal distribution of truthful and subjects. deceptive subjects. The 100 sets of charts were those used in a previous polygraph research project (Franz, 1989). Those charts were selected from the files of Argenbright, Inc., a commercial polygraph firm. The 100 sets were originally selected by taking from the company files the first 50 confirmed truthful and the first 50 confirmed deceptive cases. All of those were also confirmed by confession. There were no sets of charts used in the study in which the original examiner considered the charts to be inconclusive. Their numerical evaluation indicated a truthful or deceptive outcome. Also, all were a zone comparison technique, and all had three polygraph charts per set.

The 40 cases were conducted by 17 different examiners, 13 males and four females. All of the examiners were graduates of American Polygraph Association accredited schools. Each was licensed in the state where the polygraph examination was conducted. The cases selected for this study were comprised of 20 females, of which 13 were deceptive and seven truthful, and 20 males, of which ten were deceptive and ten were truthful. The average age for the females was 26.3 years (deceptive 24.3, truthful 31.8) with a range of 13 to 41 years, with one unknown. The average age for the males was 30.8 years (deceptive 26.7, truthful 34.6) with a range of 20 to 62 years, with one unknown. The average age for deceptive was 25.4 years and for truthful 33.7 years. The average education for females was 11.4 years of school (10.7 for deceptive, 13 for truthful) with a range of eight years to three years of college. The average education for males was 12.6 years (13.5 for deceptive, 11.8 for truthful) with a range of six years to one year of graduate work. The average education for the truthful was 11.9 years and for deceptive 12.2 years. Thirty-three of the cases involved theft, one was rape, one was a rape victim, one involved falsification of documents, one was a statement verification, on was kickback, one was fraud, one involved covering up an automobile accident. All examinations were conducted with four-channel Lafayette polygraph instruments that recorded cardiovascular patterns, thoracic respiration, abdominal respiration, and electrodermal activity. The cardio components were all electronically enhanced. The respiration was recorded by at least one electronically enhanced component. When only one electronically enhanced component was present in the respiration the other component was the standard pneumatic

component. Thirty-one of the examinations were conducted using the zone comparison as it is taught at the U.S. Department of Defense Polygraph Institute (1991). Nine of the examinations used basically the same format but replaced symptomatics with irrelevants. Irrelevants and symptomatics do not contribute to a numerical evaluation of the polygraph charts. The zone comparison technique involves the comparison of the physiological tracing of three relevant questions with adjacent control question tracings.

The 11 examiners were required to review all 40 sets of charts. A decisions was made to use 40 sets rather than a lesser number to allow for repeated opportunity for the examiner to be faced with different criteria for evaluation. These charts were not all provided at the same time, rather the examiners were sent two sets of charts on 20 separate occasions. This was not universally true as some examiners requested more as the project neared completion. Other examiners that worked in the building where the research was conducted were allowed to come in and evaluate several sets during a one-day period as their schedule allowed. As evaluation of each quantity of charts was completed and returned to the researcher an additional quantity of charts were sent to the examiner for review and evaluation. A11 examiners reviewed the same charts but not in the same order. They had no opportunity to work together or discuss the charts with each other. The evaluation was accomplished on score sheets designed specifically for this research project.

In addition to the charts the examiners were sent a form with a checklist to be used for chart analysis. These criteria were taken from those taught at DoDPI and generally accepted as those most frequently seen in chart evaluation. The list included eight criteria in the respiration, two in the electrodermal, and seven This did not preclude the examiner from in the cardiovascular. using another criterion in his analysis of the charts as a space was provided in each physiological parameter marked "Other" (see Addendum 1). A definition of each criterion in the checklist was also sent to every examiner so there would be no misunderstanding as to the examiner's identification of a criterion and the researcher's interpretation of what the examiner selection meant (see Addendum 2). Each examiner was also sent a short biographical data form to be completed and returned to the researcher (see Addendum 3).

Examiners were instructed to score charts numerically using a seven position scale. This scale provides for a score of -1, -2, or -3 if the physiological reaction is stronger to the relevant and +1, +2, or +3 if the reaction is stronger to the control The weight assigned is generally based on magnitude of the reactions and duration of reactions as well as the number of criteria present (Greene, 1980). If the reactions to the adjacent relevant

and control questions are equal or if there is a lack of reaction to the relevant and control questions being compared, a score of zero is assigned. If a value of zero was assigned, the examiner indicated whether that was a result of a distortion or whether criteria from both relevant and control cancelled out each other, and if so what criteria was involved in the decision making This scoring was accomplished for each pairing of a process. relevant and control question in each component on all charts in a case. Examiners were informed they should first assign a numerical value using the method with which they would normally score their own charts such as comparing the stronger control against the relevant question vs. comparing the relevant against the weaker control. Then, if the format permits, rescore by switching to the alternate control scoring method. The examiner was instructed to indicate which method was performed first. This was to allow a determination of the accuracy of the procedure of scoring the relevants against strong control reactions as taught by DoDPI versus the procedure advocated by Backster of scoring relevant question reactions against the weak control response. By comparing and contrasting the two a determination could be reached as to which method of control question evaluation rendered the most valid results.

Examiners were not to make a final determination of truth or deception. This was to avoid the temptation to arrive at decisive scores when they were within a point or two of doing so, and concern about their inconclusive rate. The research was not intended to determine their accuracy at reading charts, although it was obvious to the examiners that we could do so, as their final numerical designation could easily be determined by totaling all chart scores.

At any point where an examiner made a numerical evaluation he was instructed to indicate all criteria present that contributed to his decision making process. For example, if an examiner evaluated a reaction at the first relevant question on the first chart to have a +1 in the respiration, he was to indicate whether his evaluation was a result of suppression, apnea, or any of the other six criterion, or a combination of them. This process was used for each relevant/control question pairing on each chart in the three physiological components for all forty cases. After indicating which criteria were used for evaluation the examiners returned the charts, with the numerical scoring data, to the senior author.

Data was logged in as it arrived, in a detailed breakdown that was later used to make numerous evaluations and comparisons. See Addendum 4 for the logging method. The numerical score sheets were also recorded for subsequent analyses. A typical score sheet is seen in Addendum 5. The data was tabulated with respect to

Numerical Scoring of Polygraph Charts

totals of each criterion used for evaluation and the percentage of difference among the same. The frequency of criteria present was viewed in terms of its appearance on: control vs. relevant questions, number of tests, and total numbers.

All numerical designations were recorded by one, two or three for every relevant/control pairing by case for each chart and by component for each chart. This data revealed the priority given to various physiological components by tallying the examiner scores given to those components as well as the number of times that component was assigned a score.

<u>Results</u>

The examiners were correct in their blind review of the sets of charts in 361 of 440 decisions. Excluding 69 inconclusives (15.7%), they were correct in 97.3% of their decisions of truth or deception (Table 1). For truthful cases, examiners were correct in 135 of 187 decisions. Excluding 44 inconclusives (24.1%), examiners were correct in 95.1% of their decisions on truthful cases (Table 2). For deceptive cases, examiners were correct in 226 of 253 decisions. Excluding 24 inconclusives (9.5%), the examiners were correct in 98.7% of their decisions on deceptive cases (Table 3).

There were 11,682 opportunities for examiners to assign numerical values in the 440 decisions. Of these, scores of +/-1, +/-2 or +/-3 were assigned 6,474 (55.4%) times. On the remaining occasions 5,208 (44.6%) a score of zero was assigned. Zeros were indicative of a complete absence of reaction, equal reactions to both the control and relevant questions under comparison, or a distortion in the tracing of either the control or relevant question that prohibited an evaluation from taking place. On chart one, the electrodermal received the most positive or negative scores (+/-1, +/-2, +/-3)--790, followed by the cardio, 743, and respiratory, 740 (Table 4).

On chart one the majority of positive or negative scores assigned were ones. This was true in each of the individual components as well. In the respiratory 577 (78%) of the total were ones, 153 (20.7%) were twos, and 10 (1.3%) were threes. In the electrodermal, 381 (48.2%) of the total were ones, 262 (33.2%) were twos, and 147 (18.6%) were threes. In the cardiovascular, 502 (67.6%) of the total were ones, 220 (29.6%) were twos, and 23 (2.8%) were threes.

On chart two, the cardio received the most scores, 785 compared to 713 for the electrodermal, and 645 for the respiration. On chart two in the respiratory, there were 436 (67.1%) ones, 201 (31.2%) twos, and eight (1.2%) threes. In the electrodermal, there were 346 (48.5% ones, 227 (31.8%) twos, and 140 (19.6%) threes. In the cardiovascular there were 544 (69.3%) ones, 224 (28.5%) twos, and 17 (2.2%) threes.

On chart three, the cardio again received the most scores, 781 compared to 698 on the electrodermal, and 577 on the respiration. On chart three, in the respiratory there were 409 (70.9%) ones, 150 (26%) twos, and 18 (3.1%) threes. In the electrodermal, there were 324 (46.4%) ones, 207 (29.7%) twos, and 167 (23.9%) threes. In the cardiovascular there were 570 (73%) ones, 177 (22.7%) twos, and 34 (4.3%) threes. It is important to note that the percentage of threes increased on the third chart over the two preceding charts in each of the three components.

Chart number one received the most total positive and negative scores, 2,275 (35.1%) followed by chart two, 2,143 (33.1%), then chart three, 2,056 (31.8%). The cardio received the most positive or negative scores, 2,309 (35.7%) when all the charts were totaled followed by the electrodermal, 2,201 (34%), then the respiration 1,962 (30.3%).

When scores were totaled to reflect the values of ones, twos and threes, the electrodermal received the highest numerical total on each chart followed by the cardiovascular, then the respiration. The electrodermal received numerical scores totaling 3,805 (40.4%) compared to 3,074 (32.6%) in the cardiovascular, and 2,538 (26.9%) in the respiration. The variance among charts was not significant as chart one received the highest total score, 3,269 (34.7%), followed by chart two, 3,125 (33.2%), then chart three with 3,028 (32.2%). Ones contributed to 4,089 (43.4%) of the total numerical value of scores given on all charts in all cases. Twos contributed 3,642 (38.7%), and threes 1,686 (17.9%).

When examiners were instructed to indicate criterion present in the decision making process, there were 3,984 opportunities in each component to indicate one or more of the criterion present. In the respiration, suppression was indicated to have contributed to a decision of truth or deception 958 (24.6%) times followed by a baseline change 435 (11.2%) times. Each of the other criterion were indicated as being involved in the decision making process less than 10% of the time (Table 5). It is interesting that of the 71 times hyperventilation was indicated as a reaction criterion, 59 (83.1%) were by the same examiner, and 18 (31%) of those calls about hyperventilation were wrong. Nine of the 12 remaining times hyperventilation was indicated, it was by a single examiner. For the electrodermal degree of reaction was indicated 2,318 (59.5%) times. This was significantly more than any other criterion in any component. The duration of reaction followed with 625 (16%) times. In the cardiovascular, an increase and decrease in blood pressure contributed to the decision most often with 1,416 (36.3%) times followed by a blood pressure increase

along 869 (22.3%) times. All other criterion in the cardiovascular pattern were indicated less than 10% of the time (Table 5).

Examiners were instructed to write in any other criteria that may have been involved in a decision, however, all write-ins on all components added together totaled less than .4%.

TABLE 1 All Examinations (n. 440)

Cutoff	Number	Number	Number	% Correct	% TNC
00163	COLLECT	Incorrect	<u>Inconcrustve</u>	0011000	<u> </u>
+/-6	361	10	69	97.3%	15.7%
+/-5	368	13	59	96.6%	13.4%
+/-4	382	13	45	96.7%	10.2%
+/-3	386	17	37	95.8%	8.4%
+/-2	395	24	22	94.3%	4.8%
+/-1	402	28	10	93.5%	2.3%

TABLE 2No Deception Indicated (n. 187)

Cutoff <u>Scores</u>	Number <u>Correct</u>	Number <u>Incorrect</u>	Number <u>Inconclusive</u>	% <u>Correct</u>	<u>% INC</u>
+/-6	135	7	45	95.1%	24.1%
+/-5	139	10	38	93.3%	20.3%
+/-4	146	10	31	93.6%	16.6%
+/-3	149	14	24	91.5%	12.3%
+/-2	155	19	13	89.1%	7.0%
+/-1	161	21	5	88.5%	2.7%

TABLE 3 Deception Indicated (n. 253)

Cutoff	Number	Number	Number	8	
<u>Scores</u>	<u>Correct</u>	<u>Incorrect</u>	Inconclusive	<u>Correct</u>	<u>%_INC</u>
+/-6	226	3	24	98.7%	9.5%
+/-5	229	3	21	98.7%	8.3%
+/-4	236	3	14	98.7%	5.5%
+/-3	237	3	13	98.7%	5.1%
+/-2	239	5	8	98.0%	3.2%
+/-1	241	7	5	97.2%	2.0%

TABLE 4

Number of Times 1's, 2's, or 3's Were Assigned

				Sub-				Sub-				Sub-		
		Pneumo		Total		GSR		Total		Cardio)	Total	Total	
	+/-1	+/-2	+/-3		+/-1	+/-2	+/-3		+/-1	+/-2	+/-3			
Chart 1	577	153	10	740	381	262	147	790	502	220	21	743	2273	35.1%
n. 1298														n. 3894
Chart 2	436	201	8	645	346	227	140	713	544	224	17	785	2143	33.1%
n. 1298														n. 3894
Chart 3	409	150	18	577	324	207	167	698	570	177	34	781	2056	31.8%
n. 1298														n. 3894
Total	1422	504	36	1962 30.3%	1051	696	454	2201 34.0%	1616	621	72	2309 35.7%	6472 n	. 11682
n. 3894														

Total Value of Numbers Assigned

				26.9%				40.4%				32.6%		
Total	1422	1008	108	2538	1051	1392	1362	3805	1616	1242	216	3074	9422	
Chart 3	409	300	54	763	324	414	501	1239	570	354	102	1026	3028	32.2%
Chart 2	436	402	24	862	346	454	420	1220	544	448	51	1043	3125	33.2%
Chart 1	577	306	30	918	381	524	441	1346	502	440	63	1005	3269	34.7%

Polygraph 1992, 21(4)

r

Numerical Scoring of Polygraph Charts

Table 5 Number of Occurrences

PNEUMO	(n. 3894)							
	Hyper- ventil- ation	I/E Ratio	Sup- pres- sion	Base Line Change	Base Line Loss	Apnea	Ampli- tude Change	Rhythm/ Regu- larity
Chart 1	37	73	328	182	77	79	121	112
Chart 2	20	97	340	138	58	77	90	159
Chart 3	24	67	290	113	70	109	70	112
Total	71 1.8%	237 6.7%	958 24.6%	435 11.2%	205 5.2%	265 6.8%	281 7.2%	383 9.8%

GSR (n. 3954)

•

	Degree	Duration
Chart 1	823	180
Chart 2	771	186
Chart 3	724	159
Total	2318 59.5%	625 16%

CARDIO (n. 3894)

	BP Increase & Decr.	BP In- crease	BP De- crease	Pulse In- crease	Pulse De- crease	Ampli- tude Increase	Ampli- tude Decrease
Chart 1	489	267	25	6	19	4	109
Chart 2	455	279	35	11	10	4	129
Chart 3	472	323	12	11	15	9	116
Total	1416 36.3%	869 22.3%	72 1.8%	28 .7%	44 1.1%	17 .4%	354 9.1%

References Cited

Adachi, K. & Suzuki, A. (1990). Correlation structure among response measures in detection of deception. <u>Reports of the</u> <u>National Institute of Police Science</u>, <u>43</u>(3), 13-16.

Ansley, Norman (1951). <u>Notes from the Keeler Polygraph</u> <u>Institute</u>. Unpublished volume, 405 pp.

Ansley, Norman & Furgerson, Ronald M. (1987). Pioneers in the polygraph: The Federal Bureau of Investigation in the 1930's. <u>Polygraph</u>, <u>16</u>(1), 33-52.

Arther, R.O. (1971). Breathing analysis, Parts I and II. Journal of Polygraph Studies, 5(4), 1-4 & (5), 1-4.

Backster, Cleve (1961, Aug 20). <u>Annual report on polygraph</u> technique trends. Academy for Scientific Interrogation.

Backster, Cleve (1962, Sep/Oct). Technique tips and polygraph chart interpretation. <u>Newsletter of the Academy for Scientific</u> <u>Interrogation</u>, p. 4.

Backster, Cleve (1963). New standards in polygraph chart interpretation - Do the charts speak for themselves? <u>Law and</u> <u>Order</u>, <u>11</u>(6), 67-68, 71.

Backster, Cleve (1963/1979). <u>Standardized polygraph notepack</u> <u>and technique guide: Backster zone comparison technique</u>. Backster School of Lie Detection.

Backster, Cleve (1985). <u>Backster zone comparison technique:</u> <u>Chart interpretation (summary)</u>. Handout at American Polygraph Association Seminar, Reno, NV.

Backster, Cleve (1990, Aug). <u>Backster zone comparison</u> <u>technique: Chart analysis rules</u>. Backster School of Lie Detection. Paper presented at the 25th annual seminar of the American Polygraph Association, Louisville, KY.

Barland, Gordon H. & Raskin, David C. (1974, Oct). <u>Psychopathy</u> and detection of deception in criminal suspects. Paper presented at the meeting of the Society of Psychophysiological Research, Salt Lake City, UT.

Ben-Ishai, Akiva (1962, Aug). <u>Some remarks on polygraph</u> <u>research</u>. Paper presented at the annual meeting of the American Academy of Polygraph Examiners, Chicago, IL.

Polygraph 1992, 21(4)

Benussi, Vittorio (1914). Die atmungsymtome der luge. <u>Archiv</u> <u>fuer die Gesampte Psychologie</u>, <u>31</u>, 244-273. Translated and published in <u>Polygraph</u>, <u>4</u>(1)(1975), 52-76.

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

Binswanger, L. (1908). Uber das verhalten des psychogalvanischen phanomens beim assoziationexperiment. <u>Diagnostische</u> <u>assoziationstudien</u>, <u>2</u>, 113-195. Reprinted in <u>Journal of</u> <u>Psychologie and Neurologie</u>, <u>10</u>, 11+.

Block, Eugene B. (1977). <u>Lie detectors: Their history and use</u>. New York: David McKay.

Bowling, Mellberth (1978). Comparative analysis of responses on unknown and known solution stimulation tests. <u>Polygraph</u>, <u>7</u>(4), 263-265.

Buckley, Joseph P. & Senese, Louis C. (1991). The influence of race and gender on blind polygraph chart analysis. <u>Polygraph</u>, <u>20</u>(4), 247-258.

Burtt, Harold E. (1921). The inspiration/expiration ratio during truth and falsehood. <u>Journal of Experimental Psychology</u>, <u>4</u>, 1-23.

Capps, Michael H. (1991). Accuracy of individual parameters in field polygraph studies. <u>Polygraph</u>, <u>20</u>(4), 65-69.

Capps, Michael H. (1991). Predictive value of the sacrifice relevant. <u>Polygraph</u>, <u>20(1)</u>, 1-6.

Capps, Michael H. & Ansley, Norman (1992). Analysis of federal polygraph charts by spot and chart total. <u>Polygraph</u>, <u>21</u>(2).

Capps, Michael H. & Ansley, Norman (1992). Analysis of private polygraph charts by spot and chart total. <u>Polygraph</u>, <u>21</u>(2).

Capps, Michael H. & Ansley, Norman (1992). Comparison of two scoring scales. <u>Polygraph</u>, <u>21</u>(1), 38-43.

Chappell, Matthew N. (1929). Blood pressure changes in deception. <u>Archives of Psychology</u>, <u>17</u>(1), 5-39.

Clinchard, A.E. (1949, Apr 18 - May 9). Notes from the Keeler Polygraph Institute. Crosland, Harold R. (1931). The relative efficacy of visual and two auditory methods of presenting word-stimuli in an experiment which combines the association-time technique with psychogalvanic technique. <u>University of Oregon Psychology Series</u>, <u>1</u>(3), 153-202.

Crowe, Michael J., Chimarys, Michael & Schwartz, John (1988). <u>The GOT polygraph test: Scoring and validity</u>. Poster presentation at the 96th annual convention of the American Psychological Association.

Cutrow, R.J., Parks, A., Lucas, N. & Thomas, K. (1972). The objective use of multiple physiological indices in the detection of deception. <u>Psychophysiology</u>, <u>9</u>(6), 578-588.

Darrow, Chester W. (1931). The significance of the galvanic recovery curve. <u>Psychological Bulletin</u>, <u>28</u>, 678-679.

Edel, E.C. & Jacoby, J. (1975). Examiner reliability in polygraph chart analysis: Identification of physiological responses. Journal of Applied Psychology, 60(5), 632-634.

Elaad, Eitan, Ginton, Avital & Jungman, Naom (1988). <u>Respiration line length and GSR amplitude as detection measures in</u> <u>criminal guilty knowledge test</u>. Unpublished manuscript.

Elaad, Eitan & Kleiner, Murray (1986). The stimulation test in polygraph field examinations: A case study. <u>Journal of Police</u> <u>Science and Administration</u>, <u>14</u>(4), 328-333.

Elaad, Eitan & Kleiner, Murray (1990). Effects of polygraph chart interpreter experience on psychophysiological detection of deception. Journal of Police Science and Administration, 17(2), 115-123.

Feleky, A.M. (1916). The influences of emotions on respiration. Journal of Experimental Psychology, 1, 218-241.

Fere, C. (1888). Note sur les modifications de la resistance electrique sous l'influence des excitations sens orielles et des emotions. <u>Comptes Rendus Seances et Memoires de la Societe de</u> <u>Biologie</u>, <u>40</u>, 28-33, 217-219.

Ferguson, Jr., Robert J. (1966). <u>The polygraph in private</u> <u>industry</u>. Springfield, IL: Charles C Thomas.

Franz, Mark L. (1989). <u>Technical report: Relative contributions</u> of physiological recordings to detect deception. DoD Contract MDA904-88-M-6612, Argenbright Polygraph, Inc., Atlanta, GA.

311

Polygraph 1992, 21(4)

-

Gordon, Nathan J. & Cochetti, Philip M. (1987). The horizontal scoring system. <u>Polygraph</u>, <u>16</u>(2) 116-125.

Greene, Walter H. (1980). <u>Polygraph interpretation: An</u> <u>introductory reference to numerical concepts</u>. California: Walter H. Greene.

Grimsley, Douglas L. & Yankee, William J. (1985). <u>The effect of</u> response type in a relevant/irrelevant polygraph examination. DoD Contract MDA904-84-R-2284, University of North Carolina at Charlotte and A. Madley Corporation.

Haney, Kenneth W. (1972). A comparison of control and non-control question techniques. <u>Polygraph</u>, <u>1</u>(3), 136-144.

Harney, Joseph W. (1943). The analyses of respiratory criteria in deception tests: A possible source of misinterpretation. Journal of Criminal Law and Criminology, <u>34</u>, 268-273.

Harrelson, Leonard H. (1964). <u>The Keeler technique</u>. Chicago, IL: Keeler Polygraph Institute.

Hathaway, Starke (1929). A comparative study of psychogalvanic and association time measures: A new psychogalvanic apparatus. Journal of Applied Psychology, 13, 632-646.

Hathaway, Starke & Hanscomb, C.B. (1958). The statistical evaluation of polygraph records. In V.A. Leonard (Ed.), <u>Academy</u> <u>Lectures on Lie Detection</u>, Vol. 2. Springfield, IL: Charles C Thomas.

Herbold-Wootten, H. (1982). The German Tatbestandsdiagnostik: A historical review of the beginnings of lie detection in Germany. <u>Polygraph</u>, <u>11</u>(3), 246-257.

Honts, Charles R. & Driscoll, Lawrence N. (1988). A field validity study of rank order scoring system (ROSS) in multiple issue control question tests. <u>Polygraph</u>, <u>17</u>(1), 1-16.

Hunt, W.A. & Landis, C. (1935). Word-association reaction time and the magnitude of the galvanic skin response. <u>American Journal</u> of Psychology, <u>47</u>, 143-145.

Imamura, Y., Yamashita, S., Suzuki, A. & Yamaoka, K. (1960). A study of the physiological responses in lie detection. <u>Reports of</u> the National Institute of Police Science, <u>13</u>, 355-364.

Inagaki, T. & Kira, S. (1978, Sep 20-21). <u>Mechanism and</u> <u>significance of the serrated nature of respiration tracings</u>. Paper presented at the National Convention of Police Psychology at the National Institute of Police Science. Inbau, Fred E. (1942). <u>Lie detection and interrogation</u>. Baltimore: Williams & Wilkins.

Jayne, Brian C. (1990). The relative contributions of physiological recordings to physiological detection of deception. <u>Polygraph</u>, <u>19</u>(2), 105-117.

Jensen, Carl W. (1981). Frequency of occurrence of specific reaction criteria as observed in the cardio tracing. <u>Academy</u> Journal, 4(2), 5,7.

Jones, Harold E. & Wechsler, D. (1928). Galvanometric technique in studies of association. <u>American Journal of Psychology</u>, <u>40</u>, 607-612.

Joseph, C.N. (1957). Analysis of compensatory responses and irregularities in polygraph chart interpretation. In V.A. Leonard (Ed.), <u>Academy Lectures on Lie Detection</u> (pp. 93-99). Springfield, IL: Charles C Thomas.

Kasai, T. (1965). Diagnostic efficiency of heart rate in experimental and field polygraph tests. <u>Polygraph Reports</u>, pp. 60-64.

Kawakami, T. (1968). Relationship between blood pressure and pulse wave form. <u>Polygraph Reports</u>, pp. 86-91.

Keeler, Leonarde (1931). Lie detector applications. <u>IACP</u> <u>Proceedings</u> (p. 184). Reprinted in D.G. Dilworth (Ed.) <u>Silent</u> <u>Witness</u> (pp. 151-153), Gaithersburg, MD: International Association of Chiefs of Police.

Keeler, Leonarde (1936). The detection of deception. In L. Keeler, C.W. Muehlberger, C.M. Wilson, F.E. Inbau, K. Keeler, and M.E. O'Neill, <u>Outline of Scientific Criminal Investigation</u> (pp. 42-50). Chicago, IL: Northwestern University School of Law.

Kizaki, Hisakazu & Osako, Hideki (1979, May). Breathing patterns for the polygraph test. <u>Reports of the National Institute</u> of Police Science, <u>32</u>(2), 136-142.

Koll, Michael (1979). Analysis of zone chart by various pairings of control and relevant questions. <u>Polygraph</u>, <u>8</u>(2), 154-160.

Kopang, Charles (1985, May 5). <u>The Canadian Police College</u> <u>polygraph examination procedure</u>. Paper distributed at the Delta College Polygraph Workshop, Delta College, University Center, MI.

Polygraph 1992, 21(4)

Kosugi, T. & Yamashita, S. (1978). An objective analysis of GSR records in control question technique. <u>Reports of the National</u> <u>Institute of Police Science</u>, <u>31</u>(1), 40-44.

Kubis, Joseph F. (1973). Analysis of polygraph data: Dependent and independent situations. <u>Polygraph</u>, <u>2</u>(1), 42-58; (2), 89-107.

Larson, John A. (1922). The Berkeley lie detector and other deception tests. <u>Reports of the American Bar Association</u>, <u>47</u>, 619-628.

Larson, John A. (1922). The cardio-pneumo-psychogram and its use in the study of the emotions, with practical application. Journal of Experimental Psychology, 5(5), 323-328.

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

Larson, John A. & Haney, G.W. (1932). Cardio-respiratory variations in personality studies. <u>American Journal of Psychiatry</u>, <u>11</u>(6), 1035-1081.

Lee, C.D. (1937, Aug 26). Letter to John Edgar Hoover, Director of the Federal Bureau of Investigation. In N. Ansley & R.M. Furgerson (1979). Pioneers in polygraph: The Federal Bureau of Investigation in the 1930's. <u>Polygraph</u>, <u>16</u>(1), 33-52.

Lee, Clarence D. (1943, Aug 1). <u>Instruction manual for the</u> <u>Berkeley psychograph</u>. Sacramento, CA: Lee & Sons.

Lee, Clarence D. (1953). <u>The instrumental detection of</u> <u>deception: The lie test</u>. Springfield, IL: Charles C Thomas.

Lockhart, Russell A. (1975). C.G. Jung: A forgotten psychophysiologist remembered. <u>Polygraph</u>, <u>4</u>(1), 18-32.

Lombroso, Cesare (1895). <u>L'homme criminel</u>, Vol. 1 (2nd ed.). Paris: Felix Alcan.

Lombroso, Cesare (1911). <u>Crime: Its causes and remedies</u>. Boston: Little, Brown & Company.

Lombroso-Ferrero, Gina (1911). <u>The criminal man according to</u> <u>the classification of Cesare Lombroso</u>. New York: G.P. Putman's Sons.

.

MacDonald, Arthur (1905). <u>Man and abnormal man including a</u> <u>study of children in connection with bills to establish</u> <u>laboratories under federal and state governments for the study of</u> <u>criminal, pauper and defective classes, with bibliographies</u>. U.S. Senate, 3rd Session, Document No. 187.

Magiera, Alex C. (1975). Patterns of purposeful distortion. In. N. Ansley (Ed.) <u>Legal admissibility of the polygraph</u> (pp 199-209). Springfield, IL: Charles C Thomas.

Marcuse, F.L. (1953). The lie detector and the professor. <u>Tennessee Law Review</u>, <u>22</u>, 919-921.

Marcuse, F.L. & Bitterman, M.E. (1946). Minimal cues in the peak of tension procedure for determining guilt. <u>American Journal</u> of Psychology, 59(1), 144-146.

Marston, William Moulton (1938/1989). <u>The lie detector test</u>. New York: Richard R. Smith. Reprinted by the American Polygraph Association.

Matte, James Allan (1980). <u>The art and science of the polygraph</u> <u>technique</u>. Springfield, IL: Charles C Thomas.

Matte, James Allan & Reuss, Ronald M. (1989). A field validation study of the quadri-zone comparison technique. <u>Polygraph</u>, <u>18</u>(4), 187-202.

Matte, James Allan & Reuss, Ronald M. (1992). A study of the relative effectiveness of physiological data in field polygraph examinations. <u>Polygraph</u>, <u>21</u>(1), 1-22.

Miyake, Yoichi (1978). A study of skin resistance response, photoplethysmographic vasomotor response, and eye movement as indices of lie detection. <u>Reports of the National Institute of Police Science</u>, <u>31</u>(2), 88-94.

Munsterberg, Hugo (1908). <u>On the witness stand: Essays on</u> <u>psychology and crime</u>. Garden City, NY: Doubleday, Page & Company.

Nakayama, Majoto & Yamamura, Takehiko (1990). Changes of respiration pattern to the critical question on guilty knowledge technique. <u>Polygraph</u>, <u>19</u>(3), 188-198.

Ohnishi, K., Matsuno, K., Arasuna, M. & Suzuki, A. (1976). The objective analysis of physiological indices in the field of detection of deception. <u>Reports of the National Institute of</u> <u>Police Science</u>, <u>29</u>(3), 181-188.

Polygraph 1992, 21(4)

Podlesny, John A., Raskin, David C. & Barland, Gordon H. (1976, Aug). <u>Effectiveness of techniques and physiological measures in</u> <u>the detection of deception</u>. Report No. 76-5 for National Institute of Law Enforcement and Criminal Justice [Contract 75-NI-99-0001].

Rafky, David M. & Sussman, Richard C. (1985). Polygraphic reliability and validity: Individual components and stress of issue in criminal tests. <u>Journal of Police Science and</u> <u>Administration</u>, <u>13</u>(4), 283-292. Abstract in <u>Polygraph</u>, <u>19</u>(2), 157-158.

Raskin, David C. (1979). Nevada Supreme Court rules that stipulated evidence is admissible - Background to the case. <u>Polygraph Law Reporter</u>, 2(1), 10-14.

Raskin, D.C., Barland, G.H. & Podlesny, J.A. (1978). <u>Validity</u> and reliability of detection of deception. Washington, DC: Law Enforcement Assistance Administration.

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

Reid, John E. (1947). A revised questioning technique in lie detection tests. <u>Journal of Criminal Law and Criminology</u>, <u>37</u>, 542-547. Reprinted in <u>Polygraph</u>, <u>11</u>(1), 17-21.

Reid, John E. & Inbau, Fred E. (1966). <u>Truth and deception: The</u> <u>polygraph ("lie detector") technique</u>. Baltimore: The Williams and Wilkins Company.

Robbins, N.E. & Penley, W.J. (1974). Polygraph progress study: United States Postal Service. <u>Polygraph</u>, <u>3</u>(3), 247-255.

Robbins, N.E. & Penley, W.J. (1975). Review of Polygraph charts of nondeceptive subjects. <u>Polygraph</u>, <u>4</u>(3), 199-206.

Ruckmick, Christian A. (1936). <u>The psychology of feeling and</u> <u>emotion</u>. New York: McGraw-Hill Book Company.

Ryan, R. (1989). <u>The accuracy of respiration, GSR, and</u> <u>cardiovascular polygraph responses utilizing numerical evaluation</u>. Unpublished Master's thesis, Reid College.

Shimizu, Y., Hayashi, K. & Yamaoka, K. (1979, Sep 20-21). Scoring weight in the control question chart interpretation. Paper presented at the National Convention of Police Psychology at the National Institute of Police Science. Shterzer, Gideon & Elaad, Eitan (1985). Validity of the control question test in two levels of the severity of crimes. <u>Proceedings</u> of IDENTA-'85: Anti-Terrorism; Forensic Science; Psychology in <u>Police Investigations</u> (pp. 1551-166). Jerusalem, Israel.

Slowick, Stanley M. & Buckley, Joseph P. (1975). Relative accuracy of polygraph examiner diagnosis of respiration, blood pressure and GSR recordings. <u>Journal of Police Science and</u> <u>Administration</u>, <u>3</u>(3), 305-309.

Stern, Robert M., Breen, John P.. Watanabe, Takami & Perry, Bradley S. (1981). Effect of feedback of physiological information in response to innocent associations and guilty knowledge. <u>Journal</u> of <u>Applied Psychology</u>, <u>66(6)</u>, 677-681.

Summers, Walter G. (1936). Guilt distinguished from complicity. <u>Psychological Bulletin</u>, <u>33</u>, 787.

Summers, Walter G., Rev., S.J. (1936). A recording psychogalvanometer. <u>Bulletin of the American Association of Jesuit</u> <u>Scientists</u>, <u>14</u>(2), 50-56.

Summers, Walter G. (1937). A new psychogalvanometric technique in criminal investigation. <u>Psychological Bulletin</u>, <u>34</u>, 551-552.

Summers, Walter G. (1938). The electric pathometer. In <u>Proceedings of the International Association of Chiefs of Police</u> (p. 142). Washington, DC: IACP.

Summers, Walter G. (1939). Science can get the confession. Fordham Law Review, 8, 334-354.

Suzuki, Akihiro & Hikita, Yoshio (1964). An analysis of responses on polygraph: A diminution of response. <u>Reports of the National Institute of Police Science</u>, <u>17</u>, 290-295. Translated and reprinted in <u>Polygraph</u>, <u>10</u>(1), 1-7.

Suzuki, A., Izumi, S., Yamaguchi, K., Ueno A., Horigucki, K. (1977). An attempt of quantitative scoring of respiration by means of an integrator. <u>Reports of the National Institute of Police</u> <u>Science</u>, <u>30</u>, 33-39.

Suzuki, A., Ohnishi, K., Matsuno, K. & Arasuna, M. (1974). Amplitude rank score analysis of GSR in the detection of deception: Detection rates under various examination conditions. <u>Reports of the National Institute of Police Science</u>, <u>27</u>, 142-148. Translated and reprinted in <u>Polygraph</u>, <u>8(3)</u>, 242-252.

317

Polygraph 1992, 21(4)

Suzuki, A. & Watanabe, S. (1972). Systolic blood pressure changes in terms of detection of deception. <u>Reports of the</u> <u>National Institute of Police Science</u>, <u>25</u>, 145-150.

Suzuki, A., Watanabe, S., Ohnishi, K., Matsuno, K. & Arasuna, M. (1973). The objective analysis of GSR amplitudes in terms of rank scores. <u>Reports of the National Institute of Police Science</u>, <u>26</u>, 234-245. Translated and reprinted in <u>Polygraph</u>, <u>8</u>(1), 53-63.

Texas A&M University (1984). <u>Seventy-first polygraph examiners</u> <u>school</u> [notebook].

Thackray, Richard I. & Orne, Martin T. (1968). A comparison of physiological indices in detection of deception. <u>Psychophysiology</u>, 4(3), 329-339.

Thackray, R.I. & Orne, Martin T. (1968). Effects of the type of stimulus employed and the level of subject awareness on the detection of deception. Journal of Applied Psychology, 52, 234-239.

Thomas, Richard (1980). <u>An analysis of objectively evaluating</u> <u>polygrams by two examiners from two separate schools</u>. Unpublished manuscript.

Timm, Howard W. (1982). Analyzing deception from respiration patterns. <u>Journal of Police Science and Administration</u>, <u>10(1)</u>, 47-51.

Timm, Howard W. (1989). Methodological considerations affecting the utility of incorporating innocent subjects into the design of guilty knowledge polygraph experiments. <u>Polygraph</u>, <u>18</u>(3), 143-157.

Trovillo, Paul V. (1942). Some indices of deception for the interpretation of polygraph (lie detector) tests. <u>Psychological</u> <u>Bulletin</u>, <u>39</u>, 599-600.

Trovillo, Paul V. (1943). Deception test criteria: How one can determine truth and falsehood from polygraph records. <u>Journal of Criminal Law and Criminology</u>, <u>33</u>, 338-358.

Trovillo, Paul V. (1953). The polygraph needs live situation research: A reply to Professor Marcuse. <u>Tennessee Law Review</u>, <u>22</u>, 922-923.

United States Department of Defense Polygraph Institute (1991). Chart interpretation summary sheet (IN330).

United States Department of Defense Polygraph Institute (1991). Zone Comparison Lesson Plan. vanHerk, M. (1990). Numerical evaluation: Seven point scale +/-6 and possible alternatives: A discussion. <u>The Newsletter of</u> <u>the Canadian Association of Police Polygraphists</u>, 7(3), 28-47. Reprinted in <u>Polygraph</u>, 20(2)(1991), 70-79.

Wada, J. & Yamaoka, K. (1976). Effect of interaction between the relevant items and the irrelevant items on the detection of deception. <u>Reports of the National Institute of Police Science</u>, <u>29</u>, 41-45.

Waid, William M., Orne, Emily C., Cook, Mary R. & Orne, Martin T. (1981). Meprobamate reduces accuracy of physiological detection of deception. <u>Science</u>, <u>212</u>, 71-73.

Wakamatsu, T. & Yoshizumi, K. (1968). Study on respiratory wave during polygraph examination. <u>Reports of the National Institute of</u> <u>Police Science</u>, <u>21</u>, 150-155.

Watanabe, Shoichi & Kajitani, Mantaro (1973). The heart rate response as an index of the lie detection test: A study of heart rate response on field testing situations. <u>Reports of the National</u> <u>Institute of Police Science</u>, <u>26</u>(3), 191-196.

Watanabe, S. & Suzuki, A. (1977). The heart rate as an index of the lie detecting test. <u>Polygraph</u>, <u>6</u>(2), 103-112.

Watanabe, S., Yamaoka, K., Takeno, Y., Kosugi, T. & Kasuya, T. (1981). Finger vasomotor and heart rate measures as indices in detection of deception. <u>Reports of the National Institute of</u> <u>Police Science</u>, <u>34</u>(3), 29-34.

Weaver, Richard S. (1980). The numerical evaluation of polygraph charts: Evolution and comparison of three major systems. <u>Polygraph</u>, <u>9</u>, 94-108.

Weaver, Richard S. (1985). Effects of differing numerical chart evaluation systems on polygraph examination results. <u>Polygraph</u>, <u>14(1)</u>, 34-42.

Winter, John E. (1936). A comparison of the cardio-pneumopsychograph and association methods in the detection of lying in cases of theft among college students. <u>Journal of Applied</u> <u>Psychology</u>, <u>20</u>, 243-248.

Yamamura, T. (1983). Numerical scoring systems of control question technique on psychophysiological detection of deception. <u>Reports of the National Research Institute of Police Science</u>, <u>36</u>, 24-32 [text in Japanese; English abstract].

Polygraph 1992, 21(4)

Numerical Scoring of Polygraph Charts

Yamamura, Takehiko & Sasaki, Norihiko (1990). Heart rate as an index of psychophysiological detection of deception in the guilty person paradigm using the paired control question technique in field population. Japanese Journal of Physiological Psychology and Psychophysiology, 8(2), 61-69.

Yamaoka, K. (1976). A physiological study for determinants of detection of deception. <u>Bulletin of the Tokya Medical-Dental</u> <u>University</u>, <u>23</u>(1), 11-22.

Yamaoka, K. & Suzuki, A. (1973). A study of blood flow as an index of lie detection. <u>Reports of the National Institute of Police Science</u>, <u>26</u>, 206-209.

Yamaoka, K. & Suzuki, A. (1973). A study of physiological measurements as indices of lie detection. <u>Reports of the National</u> <u>Institute of Police Science</u>, <u>26</u>, 185-190.

Yankee, William J. & Grimsley, Douglas L. (1986). <u>The effect of</u> <u>a prior polygraph test on a subsequent polygraph test</u>. Contract MDA904-84-C-4249, A. Madley Corporation and the University of North Carolina at Charlotte. Abstract in <u>Psychophysiology</u>, <u>24</u>(5), 621-622.

NUMERICAL SCORING OF POLYGRAPH CHARTS: WHAT EXAMINERS REALLY DO

By

Michael H. Capps and Norman Ansley

We know something about what examiners are taught to do in reading charts of the type that may be subjected to numerical analysis. We also know something about global, rank order, statistical, computerized, and non-standard methods. There is a considerable body of literature on the results of independent numerical scoring of sets of charts from real crimes and mock crimes where the truth or deception is definitely known. Although numerical scoring is used with most control question tests, there is a difference in the way people apply numerical scoring, partly because of diverse rules, and partly from the fact that arbitrary scoring rules cannot be uniformly applied to the many variations that appear in physiological tracings. There is some literature that suggests that for those test formats that lend themselves to numerical scoring, the use of numerical scoring produces a higher degree of accuracy than global methods, but there are also some contrary findings. The history of reading the physiological recordings has not been one of unified rigorous scientific Rather, it appears that most of the current concepts inquiry. came from the codified observations of experienced polygraph examiners.

What this study sought was information on what criteria polygraph examiners currently apply when they analyze sets of control question polygraph charts. We wanted to know what they did when they were correct, what they did when they were wrong, and what was the difference. We were interested in what difference existed when examiners scored against the weakest control question response compared to the strongest control question response, a fundamental difference between two widely used scoring methods with zone comparison charts. The Backster system scores against the weakest, the DoDPI and other systems score against the strongest, and at one time the Utah system just scored the control preceding the relevant (Raskin, 1979; Weaver, 1985), but that may have changed. We wondered if some deception criteria are being taught that are never used, and that related reactions may rarely or never occur.

The senior author is a past president of the APA and Life Member who has been a regular contributor to the journal. The junior author is a Life Member of the APA and the Editor-in-Chief of APA Publications. For reprints write to P.O. Box 794, Severna Park, MD 21146.

ANOMALIES: THE CONTRIBUTIONS OF THE CARDIO, PNEUMO, AND ELECTRODERMAL MEASURES TOWARDS A VALID CONCLUSION

By

Michael H. Capps and Norman Ansley

Background

Despite the fact that scoring anomalies and errors are common there has been very little interest in them. In the early days of scoring zone comparison charts, Cleve Backster recognized them as a problem, and his scoring system required the examiner to remove from consideration the channel that went away from the trend. In those early years, the notepack required a score for each channel, followed by a notation "eliminate one" (Backster, 1963). Later, Backster revised his notepack instruction and removed the rule deleting the score for one channel (Backster, 1979). While Backster's method probably removed most anomalies, it didn't remove them all, and in the process may have given up more information than misinformation; but in reality, we really don't know if that was so.

An anomaly in testing occurs in truthful cases where the test procedure produces indications of guilt, or in deceptive cases the test procedure produces indications of truthfulness. An error in scoring is when the test result was correct, but the examiner misread the data. If these contrary indications occur too often, there is an inconclusive decision, and if they occur with exceptional frequency, there is a false positive or false negative result. Occasional occurrence is usually insufficient to alter the decision. The smallest anomaly of interest in zone comparison tests is a numerical score of +/-1 in the wrong direction in one channel at one decision spot. With a truthful person we would expect all spot scores to be zero or +1, +2, or +3. A score of -1, -2, or -3 in the test of a truthful person is misleading. In a typical zone comparison test there are nine spots, three on each chart, and there are three parts to each of those nine spots, one each for respiration, electrodermal, and cardiovascular recordings.

The study of anomalies is a vital part of the preparation for developing sound algorithms to analyze charts. In one of the contracts let to prepare for computerized chart analysis, Brian Jayne (1990) recorded the frequency and location of each anomaly,

The senior author is a past president of the APA and Life Member who has been a regular contributor to the journal. The junior author is a Life Member of the APA and the Editor-in-Chief of APA Publications. For reprints write to P.O. Box 794, Severna Park, MD 21146.

Anomalies

then prepared an analysis of the data. His data came from 100 verified control question polygraph tests, 50 truthful and 50 deceptive. In reviewing anomalous pairings resulting in a false negative or false positive spot totals, he found significantly more anomalies in electrodermal and cardiovascular records for truthful subjects than for deceptive subjects. However, these anomalies did not adversely affect the overall scores.

Similarly, Suzuki and Hikita (1964) investigated anomalies in a five question peak of tension format conducted on mock crime cases. They found that reactions that led to an erroneous interpretation occurred 38 times (4.2%) in the respiration, 56 times (6.2%) in the electrodermal, and 18 times (2%) in the cardiovascular.

There is a considerable variance in the findings of researchers who have attempted to evaluate the contributions of the data in each physiological channel. Some of the research reports evaluate the channels only by their contributions to the entire set, some break that down by chart, and one does so by spot. There are also a few reports that consider the difference in truth or deception, or by the gender of the subject. We have also included reports on the validity of specific charts. There is a paucity of research on the anomalous and erroneous scores found within chart evaluations. While this study addresses all of these topics it probably will not be definitive, even with 40 cases evaluated by 11 examiners. Nonetheless, the data in this study is from field cases, an should contribute more to our knowledge of these topics. These findings are of current value in constructing an algorithm to analyze charts as to the probability of truth or deception (Olsen, Ansley, Feldberg, Harris & Cristion, 1991).

There is an abundance of studies of field polygraph charts in which the authors have analyzed the accuracy of the cardio, pneumo, and electrodermal recordings (Capps, 1991). One study displayed the consistency with which ten examiners determined there was or was not a reaction, a study that employed 40 sets of relevant/irrelevant charts. (The study was reported twice, with the second report being the more accurate and complete, Edel & Jacoby, 1975; Edel & Moore, 1984). Based on 7,590 decisions, the raters were in agreement on 96% of the cardiovascular tracings, 91% of the electrodermal tracings, and 96% of the respiration channels. The overall agreement was 94%.

Validity of Individual Charts

A few studies have given the validity of individual charts in the sequence. Kirby (1981) reported on the validity achieved by ten examiners who scored the first chart and third chart of Reid

Control Question Test in which half of the cases were conducted with the Reid Stimulus method and half were conducted with an open stimulus method. In both methods the stimulus chart was obtained after the first relevant chart. Reading the first chart with the aid of the Reid stimulus chart, they were correct on 45%, wrong on 37%, and inconclusive on 18%. With the open stimulus method they were correct in 52%, incorrect in 26%, and inconclusive in 22%. On the third chart, using the Reid stimulus method, they were correct in 59%, incorrect in 18%, and inconclusive on 23%. With the open stimulus method they were correct on 53%, incorrect on 30%, and inconclusive on 17%.

Validity of Individual Charts

	Reid	1 Stimulu	ls	Open Stimulus				
	Correct	Error	Incon.	Correct	Error	Incon.		
First	45%	37%	18%	52%	26%	22%		
Chart	n. 90	n. 74	n. 36	n. 104	n. 52	n. 44		
Third	59%	18%	23%	53%	30%	17%		
Chart	n. 118	n. 36	n. 46	n. 106	n. 60	n. 34		

Validity of Individual Charts Inconclusives Excluded

	Reid St	imulus	Open Stimulus			
	Correct	Error	Correct	Error		
First	55%	45%	67%	33%		
Chart	n. 90	n. 74	n. 104	n. 52		
Third	77%	23%	64%	36%		
Chart	n. 118	n. 36	n. 106	n. 60		

Kleinmuntz and Szucko (1984) reported on the accuracy of three experienced examiners and three student examiners who saw one chart each from 50 truthful and 50 deceptive sets of charts. It may be that the experienced examiners saw less than the 100 charts. No inconclusive calls were allowed, and it is not known which chart students and examiners saw from each set of charts nor is it known if the authors used the same charts throughout (Buckley, 1987). There are so many deficiencies in the study that the results are of very limited value. The authors published four reports on this study, one in 1982, and three in 1984. The data they report varies somewhat, and the best understanding of the ceport seems to reflect 74% correct (26% incorrect) for the leceptive charts and 63% correct (37% incorrect) for the truthful.

Anomalies

Rafky and Sussman (1985) had four experienced examiners numerically score only the second chart of 60 sets of confirmed charts, of which 30 were deceptive and 30 were truthful. On the first viewing they saw three separate parts of each chart, the pneumo, cardio and electrodermal, scoring them independently. Six weeks later they saw all three tracings at one time on the charts and scored them again. Inconclusives deleted, these are the results:

Separately Scored Channels, Blind to the Others, Second Chart

	Pneumo		Cardio		Elect.		Combined	
Correct		93%		99%	:	100%		97%
Truthful (n. 120)	89	of 96	101	of 102	102	of 102	292	of 300
Correct		96%		99%		91%		95%
Deceptive (n. 120)	99	of 103	104	of 105	91	of 100	294	of 308

Separately Scored Channels, Others in Sight, Second Chart

	Pneumo	Cardio	Elect.	Combined
Correct	91%	97%	99%	96%
Truthful (n. 120)	90 of 99	102 of 105	100 of 101	292 of 305
Correct	96%	98%	95%	97%
Deceptive (n. 120)	100 of 104	108 of 110	96 of 101	304 of 315

Validity of Each Channel

Several authors have given estimates of the validity or the contribution of each channel toward a decision, but have not separated the data as to truth and deception or by gender (Barland & Raskin, 1974; Franz, 1989; Jayne, 1990; Ryan, 1989, Slowik & Buckley, 1975; an Suzuki, 1975).

Barland and Raskin (1974) used all the criminal cases conducted by Barland in the years 1971 through 1974 who confessed (n. 20) or pleaded guilty without plea bargaining (n. 7) as confirmed deceptive cases. There were no truthful subjects in this research.

	Pneumo	Cardio	Elect.	Overall Avg.
Correct	67%	63%	93%	74%
	18 of 27	17 of 27	25 of 27	(60 of 81)
Errors	22%	22%	4%	16%
	6 of 27	6 of 27	1 of 27	(13 of 81)
Inconclusive	11%	15%	4%	10%
	3 of 27	4 of 27	1 of 27	(8 of 81)

Accuracy by Channel of Deceptive Subjects (n. 27)

Accuracy by Channel of Deceptive Subjects Inconclusives Excluded

	Pneumo	Cardio	Elect.	Overall Avg.
	75%	74%	96%	82%
:orrect	18 of 24	17 of 23	25 of 26	(60 of 73)
	25%	26%	48	18%
rrors	6 of 24	6 of 23	1 of 26	(13 of 73)

The scores above were completed with a seven-position scale ind a +/-5 cutoff. Recomputed with a +/-1 cutoff all the decisions inconclusive results.

Franz (1989) used a computer analysis of 100 confirmed specific issue polygraph cases involving criminal activity. His inalysis with respect to the contributions of each channel was 'espiration 30% (16.7% thoracic and 13.7% abdominal), 66.5% electrodermal, and 3.5% cardio.

Ryan (1989) had four examiners numerically score 40 sets of erified specific issue tests, 20 deceptive and 20 truthful. The xaminers read the charts one channel at a time with the other hannels covered. The inconclusive rates by channel were espiration 40.8%, cardio 49.2%, and electrodermal 43.8%. Accuracy y channel is shown on the following table:

Accuracy by Channel

	Respiration	Cardio	Electrodermal
Correct	71%	69%	88%
Error	29%	31%	12%

Anomalies

Slowik and Buckley (1975) used 30 verified sets of field polygraph charts, 15 truthful and 15 deceptive. Those charts contained 141 relevant questions of which 71 were verified as truthful and 70 verified as deceptive. Seven staff examiners of John E. Reid and Associates used numerical analysis to score the charts, although they were blind to the questions asked and case They were to make decisions on the veracity of the subjects facts. to each of the 141 relevant questions then judge their overall veracity from the three channels. They did this four times at three month intervals, and each time they first saw only one channel at a time on a chart, as the other two channels were Then they saw all three channels. The seven examiners masked. made 141 calls on four occasions at 3,948 spots, on each of three channels, for a total of 11,944 decisions. In their overall decisions, using respiration they were correction 80.5%, using blood pressure they were correct in 77.2%, and using electrodermal they were correct in 80%. By two to one, the errors were false negatives over false positives. Using all three parameters they were correct in 87.2% of the cases. With individual question accuracy by channel, it was 77.5% with respiration, 72.9% with cardio, and 73.5 % with electrodermal and combined 81%. The few inconclusives were not included in the computations.

For an idea of the variability of the examiner decisions, the table below lists their accuracy by parameter. Each examiner had made a decision o the 141 relevant questions on four occasions, each three months apart. After each parameter was seen separately, examiners read all three parameters at once.

Overall Accuracy by Channel

Examiner	Respiration	Cardio	Electrodermal	All Three
1	76.7%	76.7%	76.7%	90.0%
2	80.0%	86.7%	93.4%	83.4%
3	86.7%	70.0%	66.7%	100%
4	83.4%	80.0%	86.7%	97.0%
5	73.4%	7607%	86.7%	70.0%
6	83.4%	73.4%	76.7%	86.7%
7	80.0%	76.7%	73.4%	83.4%
Average	80.5%	77.2%	80.0%	87.2%
-				

Validity of Each Channel with Truthful and Deceptive Subjects

Several authors conducted the analysis of the value of each channel but added data which compared the values for truthful subjects with deceptive subjects (Elaad, 1985; Elaad, Ginton & Jungman, 1988; Elaad & Kleiner, 1990; Slowik & Buckley, 1975; and Winter, 1936).

Elaad (1985) took 60 sets of verified control question charts from police files in Israel, 30 truthful and 30 deceptive. He computed the Receiving Operating Characteristic (ROC) for each channel in each case. He also computed the mean scores from numerical analysis of the charts. In regard to ROC, he reported that the skin resistance was more effective with the guilty than the innocent, and the cardio was the least effective with both groups. ROC is an index of the diagnostic value of a test. These are the ROC values by channel for the 60 sets of charts:

ROC Values

Respiration: .863 Cardio: .722 Electrodermal: .769

For the numerical analysis, Elaad recorded the mean scores by truth and deceptive status:

Mean Scores of Numerical Evaluations

	Respiration	Cardio	Electrodermal	Average
Truthful	+8.3	+2.5	+4.4	+5.1
Deceptive	-8.9	-2.4	-8.9	-6.73

Elaad, Ginton and Jungman (1988) used 40 sets of verified Guilty Knowledge Test (GKT) charts from police files in Israel, 20 truthful and 20 deceptive. All but one of these 40 sets were conducted after a Reid Control Question Test. Electrodermal amplitude was measured in mm. by hand and a computer measured line length, as developed by Timm (1982). Those were the only two measures used.

Accuracy of Two Limited Measures on GKT Charts

	Respiration Line Length	Electrodermal Amplitude	Combined
Truthful	95%	95%	90%
(n. 20)	19 of 20	19 of 20	18 of 20
Deceptive	50%	35%	65%
(n. 20)	10 of 20	7 of 20	13 of 20

Elaad and Kleiner (1990) used 50 sets of confirmed control question test polygraph charts, 25 truthful, 25 deceptive. Ten examiners independently scored ten sets of charts within the seven-position scale. Half the examiners were experienced and half were examiner trainees. The report said the experienced examiners had significantly better detection rates with the respiration, but there was no significant difference between the groups in scoring

Anomalies

electrodermal or cardio patterns. Using the mean scores of the two groups, the following was the accuracy of each channel:

Mean Scores of Experienced Examiners

	Respiration	Cardio	Electroderma	
Truthful	+3.8	+1.5	+0.8	
Deceptive	-3.0	-0.8	-2.9	

Mean Scores of Examiner-Trainees

	Respiration	Cardio	Electrodermal
Truthful	+2.5	+1.8	+1.8
Deceptive	-1.2	-0.3	-2.2

Elaad and Kleiner also computed the scoring accuracy for the experienced and inexperienced group by channel. The cutoff was reduced from +/-6 to +/-2 since it was by channel. The accuracy, by channel, was:

Accuracy by Channel, Experienced Examiners Inconclusives Excluded

	Respiration	Cardio	Electrodermal
Truthful	100%	73%	62%
(n. 50)	33 of 33	19 of 26	13 of 21
Deceptive	84%	70%	91%
(n. 50)	26 of 31	16 of 23	29 of 32

Accuracy by Channel, Examiner-Trainees Inconclusives Excluded

	Respiration	Cardio	Electrodermal
Truthful	96%	74%	71%
(n. 50)	25 of 26	20 of 27	17 of 24
Deceptive	81%	61%	89%
(n. 50)	17 of 21	14 of 23	24 of 27

Winter (1936) set out to compare the cardio-pneumogram with the word association/reaction time technique. The subjects were 25 women, all residents of dormitories at West Virginia University. One among them was suspected of committing thefts from the rooms. (The word association test missed the deceptive person who was in the inconclusive range and cleared only 19 of the 24 innocents.)

	Respiration	Cardio
Truthful	100%	92%
(n. 24)	24 of 24	22 of 24
Deceptive	0	100%
(n. 1)	0 of 1	1 of 1
Inconclusive	4% 1 of 25	8% 2 of 25

Winter scored his respiration and cardio channels with a simple numerical system: '0' for no significance; '1' for some significance, points in the direction of guilt; and '2' for distinct signs of guilt. Those scoring '1' were considered inconclusive. His techniques involved a "normal or practice curve" test used for comparison, and three real examinations. Case Number 11, who confessed, showed "distinct signs of guilt" at level '2' in the first and third charts and level '1' in the second chart. While this data may be of limited value because of the low base rate of deception and the lack of an electrodermal measure, it is the first research to report the use of a numerical system for evaluating respiratory and cardiovascular reactions.

Jayne (1990) used 100 sets of confirmed specific issue polygraph examinations, and had reactions measured to the nearest millimeter separately by three polygraph examiners. The research was developing background information for a computer analysis program. They also performed a standard numerical scoring of each channel and of each set of charts, and the data below are from that numerical scoring.

Excluding inconclusive opinions, numerical scoring produced an average accuracy of 92%. Averaging the examiners' quantitative results, and using the median as a cutoff score, the respiration parameter accuracy was 84%, the cardio 77%, and the electrodermal 65%. The differences were statistically significant. There is, however, no inconclusive range allowed by that method. Using the more traditional method of scoring with cutoff scores for inconclusives, the results do not provide the significant differences between truthful and deceptive people.

Anomalies

Accuracy by Channel of Truthful Subjects

	Respiration	Cardio	Electrodermal	Combined
Correct	75%	70%	69%	71%
Error	12%	15%	18%	15%
Inconclusive	13%	15%	13%	14%

Truthful Subject Accuracy, Inconclusives Excluded

	Respiration	Cardio	Electrodermal	Combined
Correct	86%	82%	79%	82%
Error	14%	18%	21%	18%

Accuracy by Channel of Deceptive Subjects

	Respiration	Cardio	Electrodermal	Combined
Correct	78%	70%	51%	66%
Error	11%	13%	35%	20%
Inconclusive	11%	17%	13%	14%

Deceptive Subject Accuracy, Inconclusives Excluded

	Respiration	Cardio	Electrodermal	Combined
Correct	88%	84%	59%	77%
Error	12%	16%	41%	23%

Validity of Each Channel by Gender

Buckley and Senese (1991) studied the effects of race and gender on the blind evaluation by nine examiners of 40 sets of confirmed specific issue polygraph examinations of which 20 were from black subjects and 20 were Caucasian subjects. Each of those sets of 20 polygraph charts had ten truthful and ten deceptive, and ten women and ten men. There was no statistically significant difference in the accuracy of blind analysis between the black and Caucasian sets of charts. Excluding inconclusives, the mean accuracy of these nine examiners was 89.5% for Caucasians and 90.6%

for blacks. The accuracy for women was 86.9% and for men was 93.4%, and the difference was not statistically significant. Of the numerous possible subcombinations between truthful black males, deceptive black males, truthful Caucasian males, deceptive Caucasian males, truthful black females, deceptive black females, truthful Caucasian females, and deceptive Caucasian females; only one showed a significant difference. That was a difference between reading the cardio channel on charts of truthful black females (75%) less accurately than the cardio channel of truthful black females (95%), however, there are only five persons in each subgroup. Inconclusive calls are excluded.

The accuracy of the individual parameters by race disclosed no statistically significant differences. Inconclusive decisions were deleted.

	Respiration	Cardio	Electrodermal
Caucasian	88%	85%	87%
Correct/Error	149/22	116/20	121/18
Black	89%	84%	79%
Correct/Error	150/19	116/22	112/30

The accuracy of the individual parameters by gender disclosed no statistically significant differences. Inconclusive decisions deleted.

	Respiration	Cardio	Electrodermal
Men	91%	87%	87%
Correct/Error	150/15	123/19	120/18
Women	85%	83%	79%
Correct/Error	149/26	109/23	113/30

Utility of Each Channel in Analyzing Charts

Matte and Reuss (1992) used 122 sets of specific issue confirmed polygraph charts of which 62 were from deceptive subjects and 53 were from truthful subjects. Matte and Reuss computed the most productive channels by gender, and by truthful or deceptive status. In terms of the productivity of the channels (but not accuracy), there are their overall findings:

Overall Productivity in Percentage

Respiration	Cardio	Electrodermal
43%	32%	24%

Anomalies

Productivity in Deceptive Cases by Gender

- -

. . .

	Respiration	Cardio	Electrodermal
Men	37%	46%	15%
Women	44%	39%	17%
Both	39%	44%	16%

Productivity in Truthful Cases by Gender

	Respiration	Cardio	Electrodermal
Men	67%	22%	11%
Women	38%	18%	43%
Both	47%	19%	33%

Since dual pneumograph patterns were used, they added that pneumo tracings from the abdominal area was most productive in 52% of the cases, thoracic in 16%, and equally productive in 33%. The thoracic/abdominal contrast is significant.

They also found differences in thoracic/abdominal dominance in terms of male or female subjects. Women produced predominantly thoracic patterns or patterns of equal value in thoracic and abdominal areas in 74% of their cases, and were predominantly abdominal reactors in 26% of their cases. However, 100% of the men were either predominantly abdominal or equally abdominal and thoracic responders, and none of the men were predominantly thoracic reactors. Matte and Reuss, however, do not address the accuracy of the channels.

Suzuki (1975) made a survey of 1,429 field cases conducted by the Japanese National Police from April to July 1973. Examiners were asked to do two things. First, indicate for each case which of the channels produced the predominate information for analysis. Second, they were asked to list in each case each channel that contributed something to their overall judgment.

Predominant Channel

Respiration	Cardio	Electrodermal	None
34%	17%	33%	16%
492 of 1429	237 of 1429	475 of 1429	225 of 1429

Channels that Contributed to the Decision

Respiration	Cardio	Electrodermal
72%	61%	70%

However, when Suzuki took a sample of 498 sets of charts from the 1,429, and asked the examiners to mark `+' and `++' for the contribution of channels. The results differed markedly for the cardio.

Respiration	Cardio	Electrodermal

72% 37% 70%

<u>Method</u>

From the score sheets of the blind reviewers, final scores were extracted by chart, by spot, and by component. Final chart scores represented the total of all scores assigned to the reactions on a given chart within an examination. For example, the final scores for chart of one of each examination reviewed by each examiner would be recorded. All 40 chart one scores from each of the 11 examiners were recorded. The same process was followed with chart two then chart three. After this was completed, the final spot scores were also recorded. Spot scores represented all of the scores assigned to each relevant question on the charts in each examination. This was accomplished by listing scores sequentially for spots one, two and three for each of the 40 examinations from the score sheets of each of the 11 examiners. Although spots one and two were present on every examination, two examinations did not contain a third spot. Since two of the 40 cases had only two relevant questions, this produced 440 scores each for relevant one and relevant two, but only 418 scores for relevant three. In zone comparison type examinations, relevants one and two are essentially the same question, whereas relevant three is an evidence connecting question. The final scores for each parameter were then recorded. This represented the total of all scores assigned in that parameter during the examination. The three parameters were the respiratory, electrodermal, and cardiovascular channels.

In an effort to identify anomalies and erroneous calls in the scoring process, all final negative scores were considered to be deceptive calls and all final positive scores were considered to be truthful calls.

<u>Results</u>

Chart one produced the most anomalous or incorrect calls at 65 (14%), followed by chart two which decreased to 47 anomalous or incorrect calls (11%), and chart three further decreased to 36 anomalous or incorrect calls (8%) (Table 1). On chart one, three of the 40 cases produced 28 (43%) of the 65 anomalous or incorrect calls. Similar situations occurred on chart two where three cases represented 20 (43% of the anomalous or incorrect calls, and on chart three where three cases represented 16 (44%) of those calls.
Only one case appears twice among the lists of cases contributing to the majority of anomalous or incorrect calls.

When reviewing the anomalies and errors by spot, relevant question #1 produced 38 of 440 incorrect calls (8.6%), relevant question #2 produced 37 of 440 incorrect calls (8.4%), and relevant question #3 produced the most incorrect calls, 55 of 418 (13.2%) (Table 4). In each situation, three cases contributed to approximately half of the erroneous calls. For relevant question #1, three cases contributed to 18 incorrect calls (47.4%). On relevant question #2, three cases contributed to 18 of the incorrect calls (49%) and on relevant question #3, there were three cases that contributed to 27 of the incorrect calls (49%).

When the scores were separated by component, 148 incorrect calls occurred. The respiratory channel produced the most, 65 incorrect calls (15%). The electrodermal produced 28, the fewest incorrect calls (6%), and the cardiovascular produced 55 incorrect calls (12.5%) (Table 7). Both the respiratory and cardiovascular channels contributed significantly more incorrect calls then the electrodermal. When confirmed truthful were distinguished from confirmed deceptive, the truthful produced 34 (18%) incorrect calls in the respiratory channel, 24 (13%) incorrect calls in the electrodermal channel, and 22 (12%) incorrect calls in the cardiovascular channel, for a total of 81 (14%) (Table 8). For the deceptive, there were 31 (12%) incorrect calls in the pneumo, three (1%) in the electrodermal, and 33 (13%) in the cardiovascular, for a total of 67 (9%) (Table 9). The number of incorrect calls by component for the truthful was significantly greater than for the deceptive.

These anomalies and errors in scoring were further separated by male and female. The charts of the male subjects produced anomalous scores on 20 cases (9%) in the respiration, eight (4%) in the electrodermal, and 14 (6%) in the cardiovascular, for a total of 41 (6%). There were significantly more incorrect calls in the component scores for females than for males; with 45 (20%) in respiration, 20 (9%) in electrodermal and 41 (19%) in cardiovascular, for a total of 106 (16%). The respiratory producing the most incorrect calls for both males and females, and the electrodermal producing the least for each group (Table 10).

An analytic method that was once briefly used in early administration of Backster zone comparison tests involved eliminating the tracing and/or the chart that was contrary to the trend of the other charts. This method was applied to both the component scores and chart scores. We tried applying this rule to our chart scores. When the component score that was opposite the trend of the other component scores was eliminated, the call was correct in only 69 (79%) of the 87 time these circumstances

occurred. When this same method was applied to eliminating the chart that was opposite to the trend, the call was correct in only 82 (77%) of the 106 cases where these circumstances occurred. In both situations the accuracy rate is far below that of the examiner using traditional scoring procedures that involve the use of all components and all charts.

Discussion

Chart number three produced the least incorrect calls of each of the three charts, followed by chart two, then chart one, which produced the most. This was similar to the finding in a previous study which investigated the error rate by chart (Capps & Ansley, 1992).

Table 1 - All (n. 440)

	Correct	Incorrect	Inconclusive	
Chart	1	357 (81.2%)	65 (14.8%)	18 (4.1%)
Chart	2	376 (85.5%)	47 (10.7%)	17 (3.9%)
Chart	3	373 (84.8%)	36 (8.2%)	31 (7.0%)

Table 2 - Truthful (n. 187)

		Correct	Incorrect	Inconclusive	
Chart	1	131 (70.0%)	46 (24.6%)	10 (5.3%)	
Chart	2	148 (79.1%)	30 (16.0%)	9 (4.8%)	
Chart	3	150 (80.0%)	20 (10.7%)	17 (9.1%)	

Table 3 - Deceptive (n. 253)

	Correct	Incorrect	Inconclusive	
Chart 1	227 (89.7%)	18 (7.1%)	8 (3.2%)	
Chart 2	228 (90.1%)	17 (6.7%)	8 (3.2%)	
Chart 3	223 (88.1%)	16 (6.3%)	14 (5.5%)	

Table 4 - All

		Correct	Incorrect	Inconclusive	
R1	(n.440)	372 (84.5%)	38 (8.6%)	30 (6.8%)	
R2	(n.440)	379 (86.1%)	37 (8.4%) 55 (12.2%)	24 (5.5%) 25 (8.4%)	
RS	(11.410)	320 (10.5%)	55 (IS.2%)	JJ (0.4%)	

Table 5 - Truthful (n. 187)

	Correct	Incorrect	Inconclusive
R1	153 (81.8%)	17 (9.1%)	17 (9.1%)
R2	146 (78.1%)	25 (13.4%)	16 (8.6%)
R3	128 (68.4%)	36 (19.3%)	23 (12.3%)

Table 6 - Deceptive

	Correct	Incorrect	Inconclusive	
R1 (n.253)	220 (87.0%)	21 (8.3%)	13 (5.1%)	
R2 (n.253)	233 (92.1%)	12 (4.7%)	8 (3.2%)	
R3 (n.231)	200 (86.6%)	19 (8.2%)	12 (5.2%)	

Table 7 - All (n. 440)

	Correct	Incorrect	Inconclusive	
Respiratory	318 (72.3%)	65 (14.8%)	57 (13.0%)	
Electrodermal	391 (88.9%)	28 (6.4%)	21 (4.8%)	
Cardiovascular	363 (82.5%)	55 (12.5%)	22 (5.0%)	

Table 8 - Truthful (n. 187)

	Correct	Incorrect	Inconclusive	
Respiratory Electrodermal	116 (62.0%) 145 (77.5%)	34 (18.2%) 25 (13.4%)	37 (19.8%) 17 (9.1%)	
Cardiovascular	154 (82.4%)	22 (11.8%)	11 (5.9%)	

Table 9 - Deceptive (n. 253)

	Correct	Incorrect	Inconclusive	
Respiratory	202 (79.8%)	31 (12.3%)	20 (7.9%)	
Electrodermal	246 (97.2%)	3 (1.2%)	4 (1.6%)	
Cardiovascular	209 (82.6%)	33 (13.0%)	11 (4.3%)	

Table 10 - Anomalies by Gender

	Male (n. 220)	Female (n. 220)
Respiratory	20 (9%)	45 (20%)
Electrodermal	8 (4%)	20 (9%)
Cardiovascular	14 (6%)	41 (19%)
Total	42 (6%)	106 (16%)

Discussion

In a report on all studies of the reliability of blind chart analyses from real cases conducted since 1980 (Ansley 1990), the researcher found accuracy of blind evaluators to be 89% for truthful subjects and 94% for deceptive subjects, with a total accuracy of 90%. Examiners conducting blind reviews in this study fared better for both truthful (95.1%) and deceptive (98.7%) with an overall accuracy of 97.3%. Conversations with the blind reviewers in this study indicated that being forced to identify the criteria that caused them to make the decisions to give a score or not give a score may have improved their accuracy. Additionally, this method of evaluation caused the examiners to spend from onehalf to one hour per case in chart analysis as opposed to current field practices of evaluating a set of charts in minutes. Another possible contribution to the accuracy of the examiners is the fact that they were not instructed to make a call of deception indicated or no deception indicated. They were to simply score the charts Although there was no problem in the researchers individually. extracting the final scoring data from the score sheets, the project was not designed to determine the accuracy of blind reviewers.

Of the ten errors made by the blind reviewers, four occurred on one case. Six of the remaining seven examiners called that case inconclusive. Since these examiners were extremely accurate on all of their other calls, this case deserves special attention. The case was confirmed truthful by the original examiner. It involved the theft of \$1,750 cash from the evidence locker of a sheriff's department during the month of December. Each of those persons who had access to the money during that time period were scheduled for a polygraph examination. On the morning of January 5th, one of the employees of the sheriff's department confessed to the theft of the missing money. Several hours later that same day the subject of the case under discussion took a polygraph examination. Although the examiner called the test result truthful, he took a statement from the subject wherein she admitted the theft of \$1,000 cash from the evidence locker on one occasion and \$200 cash on a second Both of these thefts occurred prior to December. occasion. She acknowledged no involvement in the December theft and there was no

reason to believe she was involved. This case is remarkable since ten of the 11 examiners saw significant physiological reactions to at least one of the relevant questions yet the subject was truthful to the offense under investigation.

Reviewing the numerical scoring of examiners we found zeros were assigned almost as often as +/- (positive or negative) scores. Although examiners were instructed to indicate whether or not the zero was due to a lack of reaction, to equal reactions at the control and relevant questions, or to distortions which prevent a comparison, they seldom explained the zero scores. It could not be determined, therefore, whether reactions occurred more frequently, the extent of distortions, or the lack of reactions. What was apparent involved the fact that the number of times positive or negative scores were assigned decreased on the second chart and again on the third chart in both the respiration and electrodermal. This was not true, however, for the cardiovascular scores. Although this trend might suggest that the control and relevant habituate at the same rate therefore producing less scores, it is important to note that scores are based on comparison rather than individual repetition and the lack of score could be from reactions of equal value regardless of whether or not they increased, decreased or remained constant.

In reviewing further the number of the positive and negative scores assigned, 63% were ones, compared to 28% twos, and 9% threes. When comparing the reactions of control versus relevant it appears that the distinction between the two is subtle significantly more often than it is obvious or dramatic. Even though the number of ones far exceeds that of twos and threes, the value each contributes demonstrates a meaningful difference.

As previously indicated, most of the evaluation criteria have been taught as they are today for over half a centurey. When examiners are actually required to evaluate charts identifying each criterion many of those taught are either not used in evaluation or not seen by examiners as contributing to the correct identification of reactions.

<u>References Cited</u>

Backster, Cleve (1963/1979). <u>Standardized polygraph notepack</u> and technique guide: Backster zone comparison technique. Backster School of Lie Detection.

Barland, Gordon H. & Raskin, David C. (1974, Oct). <u>Psychopathy</u> and detection of deception in criminal suspects. Paper presented at the meeting of the Society of Psychophysiological Research, Salt Lake City, UT. Buckley, Joseph P. (1987). Buckley refutes assertion in AMA Journal. <u>APA Newsletter</u>, <u>20</u>(4), 9-13.

Buckley, Joseph P. & Senese, Louis C. (1991). The influence of race and gender on blind polygraph chart analysis. <u>Polygraph</u>, <u>20(4)</u>, 247-258.

Capps, Michael H. (1991). Accuracy of individual parameters in field polygraph studies. <u>Polygraph</u>, <u>20</u>(4), 65-69.

Capps, Michael H. & Ansley, Norman (1992). Analysis of federal polygraph charts by spot and chart total. <u>Polygraph</u>, <u>21</u>(2).

Capps, Michael H. & Ansley, Norman (1992). Analysis of private polygraph charts by spot and chart total. <u>Polygraph</u>, <u>21</u>(2).

Edel, E.C. & Jacoby, J. (1975). Examiner reliability in polygraph chart analysis: Identification of physiological responses. Journal of Applied Psychology, 60(5), 632-634.

Edel, E.C. & Moore, Jr., L.A. (1984). Analysis of agreement in polygraph charts. <u>Polygraph</u>, <u>13</u>(2), 193-197.

Elaad, Eitan (1985). <u>Validity of the control question test in</u> <u>criminal cases</u>. Unpublished manuscript, Israel Police Headquarters, Jerusalem, Israel.

Elaad, Eitan, Ginton, Avital & Jungman, Naom (1988). <u>Respiration line length and GSR amplitude as detection measures in</u> <u>criminal quilty knowledge test</u>. Unpublished manuscript.

Elaad, Eitan & Kleiner, Murray (1990). Effects of polygraph chart interpreter experience on psychophysiological detection of deception. Journal of Police Science and Administration, 17(2), 115-123.

Franz, Mark L. (1989). <u>Technical report: Relative</u> <u>contributions of physiological recordings to detect deception</u>. DoD Contract MDA904-88-M-6612, Argenbright Polygraph, Inc., Atlanta, GA.

Jayne, Brian C. (1990). The relative contributions of physiological recordings to physiological detection of deception. <u>Polygraph</u>, <u>19</u>(2), 105-117.

Kirby, Steven L. (1981). The comparison of two stimulation tests and their effect on polygraph technique. <u>Polygraph</u>, <u>10(2)</u>, 63-76.

Kleinmuntz, Benjamin & Szucko, Julian J. (1984). On the fallibility of lie detection. <u>Law & Society Review</u>, <u>17</u>(1), 85-104.

Kleinmuntz, Benjamin & Szucko, Julian J. (1984). A field study of the fallibility of polygraph lie detection. <u>Nature</u>, <u>308</u>(5958), 449-450.

Matte, James Allan & Reuss, Ronald M. (1992). A study of the relative effectiveness of physiological data in field polygraph examinations. <u>Polygraph</u>, <u>21</u>(1), 1-22.

Olsen, D.E., Ansley, N., Feldberg, I.E., Harris, J.C. & Cristion, J.A. (1991). Recent developments in polygraph technology. Johns Hopkins APL Technical Digest, 12(4), 346-356.

Rafky, David M. & Sussman, Richard C. (1985). Polygraphic reliability and validity: Individual components and stress of issue in criminal tests. Journal of Police Science and Administration, 13(4), 283-292. Abstract in Polygraph, 19(2), 157-158.

Ryan, R. (1989). <u>The accuracy of respiration, GSR, and</u> <u>cardiovascular polygraph responses utilizing numerical evaluation</u>. Unpublished Master's thesis, Reid College.

Slowick, Stanley M. & Buckley, Joseph P. (1975). Relative accuracy of polygraph examiner diagnosis of respiration, blood pressure and GSR recordings. <u>Journal of Police Science and</u> <u>Administration</u>, <u>3</u>(3), 305-309.

Suzuki, Akihiro (1975). Field polygraph condition and analysis of its effective procedures. <u>Reports of the National Institute of</u> <u>Police Science</u>, <u>28</u>, 15-22.

Suzuki, Akihiro & Hikita, Yoshio (1964). An analysis of responses on polygraph: A diminution of response. <u>Reports of the National Institute of Police Science</u>, <u>17</u>, 290-295. Translated and reprinted in <u>Polygraph</u>, <u>10</u>(1), 1-7.

Winter, John E. (1936). A comparison of the cardio-pneumopsychograph and association methods in the detection of lying in cases of theft among college students. <u>Journal of Applied</u> <u>Psychology</u>, <u>20</u>, 243-248.

340

STRONG CONTROL VERSUS WEAK CONTROL

By

Michael H. Capps and Norman Ansley

Background

For over 30 years there has been controversy over procedures used to analyze zone comparison polygraph charts. Backster, who developed the zone comparison technique in 1960, has theorized that the comparison between control and relevant questions should involve the adjacent control question with the lesser response. (Backster, 1990). This was only true unless another adjacent control contained a reaction four (4) times greater than the relevant to which it was juxtaposed. In this case the control with the greater reaction could not be ignored for purposes of comparisons. Further, if a lack of reaction occurred at the relevant question, the relevant must be compared to the adjacent control with the greater reaction (Weaver, 1980).

In the past, the U.S. Army School has taught that relevant question '5' (the first relevant) was compared with either control question '4' or control question '6,' whichever displayed the greatest reaction in degree, duration and/or number of reaction criteria. Relevant question '7' was compared to control question '6' and relevant question '10' was compared to control question '9' (Greene, 1980). The Army Criminal Investigation Division Command (CIDC) policy varied somewhat in that if a reaction to symptomatic question '8' was equal to or greater than control question '6' and relevant question '7,' no evaluation was made at that spot (Koll, 1979). Beginning in 1982 or 1983, the U.S. Army CIDC instituted a policy that changed the previous guidelines involving chart interpretation. This had to do with reactions to the symptomatic question '8.' If after evaluating each component on each chart symptomatic question '8' had a greater reaction than control question '6' and relevant question '7' more often than question '6' and '7' had to symptomatic '8', the entire test was determined to be inconclusive regardless of the overall score (Schwartz, 1992).

The Naval Investigative Service varied even more by comparing the greater of control question '6' or symptomatic '8' to relevant question '7.'

The senior author is a past president of the APA and Life Member who has been a regular contributor to the journal. The junior author is a Life Member of the APA and the Editor-in-Chief of APA Publications. For reprints write to P.O. Box 794, Severna Park, MD 21146.

The Air Force identified the strongest control on the first chart and transposed it to control question '6' position on subsequent charts where the response to both relevant questions '5' and '7' would be compared to responses at that control question spot (Koll, 1979). Raskin, a prolific author of polygraph research reports, modified the procedure of relevant-control comparison to a more simplistic method of comparing the relevant question reaction to that of the immediately preceding control (Raskin, 1979). Recent research with computerized polygraph data has calculated the mean and standard deviation for each control and relevant question After all subject responses are standardized, a parameter. comparison is made in which relevant questions demonstrate negative scores for the confirmed truthful and positive scores for the confirmed deceptive subject (Olsen, et.al., 1992).

Many of the heterogeneous systems of comparing the control to the relevant were empirically derived without a sound theoretical Some rules were developed because of errors made using basis. previously instituted general methods of analysis. Other rules emerged because they seemed logical, while few rules were developed as a result of scientific research. This study investigated the outcomes of scoring using the current method taught at the Department of Defense Polygraph Institute of comparing the stronger reaction control of question '4' or '6' against relevant '5,' control question '6' against relevant '7,' and control question '9' against relevant '10'. This was contrasted with outcomes of the Backster method of using the weaker reaction control of question '4' or '6' to relevant '5,' control question '6' against relevant '7,' and control question '9' against relevant '10.' In fairness to Backster this approach should not be taken unless all the rules that are involved in the chart interpretation technique instituted by Backster are Nevertheless these systems are being contrasted to employed. determine the accuracy based on varied cutoffs.

Procedures

Examiners were required to blind score forty sets of confirmed zone comparison charts using a seven-position scale (-3, -2, -1, 0, +1, +2, +3). When possible the examiners were to score against the strong control and separately score against the weak control. In polygraph examinations conducted using the zone comparison technique the first relevant question has a control question placed on each side. The adjacent control question that receives the greater physiological reaction in each parameter is referred to as the strong control for purposes of scoring that parameter. The control question that generates the lesser response is considered to be the weak control for purposes of scoring. The examiner compares the response of the relevant

question to the response of an adjacent control to determine the numerical designation that will be assigned. Circumstances may evolve that do not allow for the placement of the control adjacent to the relevant on both sides, such as the use of an irrelevant question immediately before the relevant question or just after the relevant question. This situation prevents the scoring of the first relevant question against two controls. In nine of the forty cases that the examiners analyzed a dual scoring option was not available. Examiner scores against both the strong and weak control were recorded by case and by examiner. The nine cases that did not have the optional scoring capability available were listed with the strong control data. This approach was taken since most examiners are trained to score against the strong control rather than the weak control (Capps & Ansley, 1992). Although examiners were not required to make a judgment of truth or deception, in this research numerical cutoffs were used by researchers for that purpose.

<u>Results</u>

Cutoffs from +/-1 to +/-6 were tallied for accuracy rates (see Table 1). Of 440 scoring decisions on cases, blind reviewers made ten errors (2.3%) with 69 inconclusive calls (15.7%) and 361 correct decisions (82%), using a +/-6 cutoff. Excluding inconclusives, the reviewers were correct in their analysis in 97.3% of the cases using the strong control procedure. Inconclusives cannot be included as errors for two reasons. The examiner did not reach a conclusion as to truth or deception on inconclusive calls therefore the examiner cannot be wrong. Secondly, other research has demonstrated that when the examiner is forced to make a call on an inconclusive result, he is correct as often as he is in error (Capps, 1991).

Analysis of the same charts by the same reviewers using the weak control scoring method yielded less accurate results. Cutoff scores from +/-1 to +/-6 were also tallied for accuracy rates (see Table 2). Of 341 decisions on cases, reviewers made 26 errors (7.6%) with 84 inconclusive calls (24.6%) and 231 correct decisions (67.7%). Excluding inconclusives, the reviewers were correct in 89.9% of the cases using the weak control scoring method.

Discussion of Strong-Weak Control Decisions

At every level of cutoff scores from +/-6 to +/-1, the percent of correct decisions is higher for the strong control method. Also, the use of the strong control produces fewer inconclusive results at +/-6 to +/-2, but at +/-1 the weak control produces 1.5% inconclusives versus 2.3% for the strong control procedure.

Raskin has commented on the issue of strong versus weak control reactions in scoring, stating that, "The scientific evidence indicates that the Backster system is inherently biased against the innocent subject and produces an unacceptable number of false positive errors" (Raskin, 1986, 1989). Weaver compared the USAMPS (Army), Backster, and Utah scoring systems. He noted that:

"The USAMPS and Utah scoring systems were not significantly different in either type of decision, numerical field score, or the difference between numerical field score and the value needed to reach a decision. However, the Backster system did differ from both the USAMPS and Utah systems. For subjects classified DI by USAMPS, Backster scores were significantly This occurs because the Backster system more negative. generally evaluates the response of a relevant question against the weakest control, unlike USAMPS or Utah ... The opposite results were obtained with the NDI and inconclusive Backster numerical scores were not significantly aroups. different ... [but] Backster scores were further from the value needed to make a decision ... The Backster scoring system produced a similar number of DI decisions, more inconclusive decision, and fewer NDI decisions than either USAMPS or Utah."

Our finding that use of the strong control is more accurate and reduces inconclusives is in concert with the findings of others, but we are not yet prepared to recommend a change in the Backster system. We do believe our findings are sufficient to justify a full study on this aspect of scoring zone comparison charts.

Discussion of Optimal Cutoff Scores

Several studies have suggested that optimal cutoff scores could be less than those of Backster (for three charts), +7 and -13, and the DoDPI at +/-6. Raskin, Barland and Podlesny in 1978 found that +/-4 was optimal. Raskin and Hare (1978) said the optimal range was +/-2 to +/-4. Shterzer and Elaad (1985) found a +/-1 compares favorably with +/-6. Elaad and Kleiner (1990) found that +/-3 gave a higher accuracy and fewer inconclusives than +/-6. The results of this study are:

TABLE 1Strong Control (n. 440)

Cutoff	Number	Number	Number	8	
<u>Scores</u>	<u>Correct</u>	<u>Incorrect</u>	<u>Inconclusive</u>	<u>Correct</u>	<u>% INC</u>
+/-6	361	10	69	97.3%	15.7%
+/-5	368	13	59	96.6%	13.4%
+/-4	382	13	45	96.7%	10.2%
+/-3	386	17	37	95.8%	8.4%
+/-2	395	24	22	94.3%	4.8%
+/-1	402	28	10	93.5%	2.3%

TABLE 2 Weak Control (n. 341)

Cutoff	Number	Number	Number	%	
<u>Scores</u>	<u>Correct</u>	<u>Incorrect</u>	<u>Inconclusive</u>	<u>Correct</u>	<u>% INC</u>
+/-6	231	26	84	89.9%	24.6%
+/-5	241	34	66	87.6%	19.3%
+/-4	249	42	50	85.6%	12.3%
+/-3	265	46	30	85.2%	8.8%
+/-2	269	55	17	83.0%	5.0%
+/-1	275	61	5	81.8%	1.5%

TABLE 3Strong NDIs (n. 187)

Cutoff <u>Scores</u>	Number <u>Correct</u>	Number <u>Incorrect</u>	Number <u>Inconclusive</u>	% <u>Correct</u>	<u>% INC</u>
+/-6	135	7	45	95.1%	24.1%
+/-5	139	10	38	93.3%	20.3%
+/-4	146	10	31	93.6%	16.6%
+/-3	149	14	24	91.5%	12.3%
+/-2	155	19	13	89.1%	7.0%
+/-1	161	21	5	88.5%	2.7%

Strong Control Versus Weak Control

TABLE 4 Strong DIs (n. 253)

Cutoff	Number	Number	Number	%	
<u>Scores</u>	<u>Correct</u>	<u>Incorrect</u>	<u>Inconclusive</u>	<u>Correct</u>	<u>% INC</u>
+/-6	226	3	24	98.7%	9.5%
+/-5	229	3	21	98.7%	8.3%
+/-4	236	3	14	98.7%	5.5%
+/-3	237	3	13	98.7%	5.1%
+/-2	239	5	8	98.0%	3.2%
+/-1	241	7	5	97.2%	2.0%

TABLE 5 Weak NDIs (n. 176)

Cutoff <u>Scores</u>	Number <u>Correct</u>	Number <u>Incorrect</u>	Number <u>Inconclusive</u>	% <u>Correct</u>	<u>% INC</u>
+/-6	71	25	80	74.0%	45.5%
+/-5	81	33	62	71.1%	35.2%
+/-4	89	41	46	68.5%	26.1%
+/-3	103	45	28	69.6%	15.9%
+/-2	107	54	15	66.5%	8.5%
+/-1	113	59	4	65.7%	2.3%

TABLE 6 Weak DIs (n. 165)

Cutoff <u>Scores</u>	Number <u>Correct</u>	Number <u>Incorrect</u>	Number <u>Inconclusive</u>	% <u>Correct</u>	<u>%_INC</u>
+/-6	160	1	4	99.4%	2.4%
+/-5	160	1	4	99.4%	2.4%
+/-4	160	1	4	99.4%	2.4%
+/-3	162	1	2	99.4%	1.2%
+/-2	162	1	2	99.4%	1.2%
+/-1	162	2	1	98.8%	.6%

<u>Cutoff</u>	<u>% Correct</u>	<u> </u>
+/-6	97.3%	15.7%
+/-5	96.6%	13.4%
+/-4	96.7%	10.2%
+/-3	95.8%	8.4%
+/-2	94.3%	4.8%
+/-1	93.5%	2.3%

TABLE 7 Overall Tradeoff by Cutoff

The data suggests that the cutoff scores an examiner or agency chooses to use may be a matter of policy. In numerical scoring of zone comparison charts, using the strong control, there appears to be a consistent relationship of inconclusive results The cutoff of +/-6 (DoDPI standard) is correct at and accuracy. 97.3%, but produces an inconclusive rate of 15.7%. To increase utility by reducing inconclusive results, there is a loss of accuracy. At a +/-1 the inconclusive rate has dropped over 13% to 2.3%, but the accuracy has also dropped 3.8% to 93.5% (Table 7). While other writers have spoken of optimum cutoff scores, one could really justify any of these cutoff scores depending on the values in the system. If reexaminations are not possible and a decision is vital, one could opt for the +/-1, accepting the lower accuracy. If reexaminations are easily arranged, the highest inconclusive rate may be less important than obtaining the highest accuracy rate.

References Cited

Backster, Cleve (1990, Aug). <u>Backster Zone Comparison Technique:</u> <u>Chart Analysis Rules</u>. Backster School of Lie Detection. Paper presented at the 25th annual seminar of the American Polygraph Association, Louisville, KY.

Capps, Michael H. (1991). Accuracy of individual parameters in field polygraph studies. <u>Polygraph</u>, <u>20</u>(4), 65-69.

Capps, Michael H. & Ansley, Norman (1992). Comparison of two scoring scales. <u>Polygraph</u>, <u>21</u>(1), 39-43.

Elaad, Eitan & Kleiner, Murray (1990). Effects of polygraph chart interpreters experience on psychophysiological detection of deception. Journal of Police Science and Administration, 17(2), 115-123.

Greene, Walter H. (1980). <u>Polygraph Interpretation: An</u> <u>Introductory Reference to Numerical Concepts</u>. California: Walter H. Greene.

Koll, Michael (1979). Analysis of zone charts by various pairings of control and relevant questions. <u>Polygraph</u>, <u>8</u>(2), 154-160.

Olsen, D.E., Ansley, N., Feldberg, I.E., Harris, J.C. & Cristion, J.A. (1991). Recent development in polygraph technology. Johns Hopkins APL Technical Digest, 12(4), 346-356.

Raskin, D.C. (1979). Nevada Supreme Court rules that stipulated evidence is admissible: Background to case. <u>Polygraph Law</u> <u>Reporter</u>, 2(1), 10-14.

Raskin, D.C. (1986). The polygraph in 1986: Scientific, professional and legal issues surrounding application and acceptance of polygraph evidence. <u>Utah Law Review</u>, <u>29(1)</u>, 29-74.

Raskin, D.C. (1989). <u>Psychological methods in criminal</u> <u>investigation and evidence</u>. New York: Springer Publishing Company.

Raskin, D.C., Barland, G.H. & Podlesny, J.A. (1978). <u>Validity</u> and reliability of detection of deception. Washington, DC: Law Enforcement Assistance Administration.

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

Schwartz, John R. (1992). Personal communication.

Shterzer, Gideon & Elaad, Eitan (1985). Validity of the control question test in two levels of the severity of crimes. <u>Proceedings</u> of the IDENTA-'85, Anti-Terrorism; Forensic Science; Psychology in <u>Police Investigations</u> (pp. 155-166). Jerusalem, Israel.

Weaver, Richard S. (1980). The numerical evaluation of polygraph charts: Evaluation and comparison of the major systems. <u>Polygraph</u>, 9(2), 94-108.

348

MACKENZIE'S EARLY POLYGRAPH INSTRUMENT

A footnote in the text <u>Truth and Deception</u> by Reid and Inbau (1977) cites an article by William O. Gay (1948) in which he states that "The Ink Polygraph" first came to our attention in an article in which the statement was made that "the polygraph is really a modification of a device invented by Sir James Mackenzie, the famous heart specialist." Reid and Inbau further noted that Mackenzie first described the instrument in an article entitled "The Ink Polygraph" which appeared in a 1908 edition of the <u>British</u> <u>Medical Journal</u>.

However, the Mackenzie instrument is by no means the first recording polygraph instrument. Several recording polygraph instruments are described and illustrated in a chapter "Psychophysical and Anthropometrical Instruments of Precision," by Arthur Macdonald in <u>The Hearings on the Bill (H.R. 14798) to Establish a</u> <u>Laboratory for the Study of the Criminal, Pauper, and Defective Classes ... before the Committee on the Judiciary, in 1902.</u> Macdonald, then in the U.S. Bureau of Education, lists instruments manufactured by Ludwig, Verdin, Marey, Mosso, Filliatre, Sommer, Von Frey, Philadelphien, and Duchene.

When James R. Wygant visited the University of Oxford's Museum of the History of Science on Broad Street in Oxford, he attempted to photograph the Mackenzie instrument. The light was too poor to obtain a good picture, so we asked the Museum for photograph and description, which they kindly furnished. W.D. Hackmann, Acting Curator, said he had not realized at the time he prepared the exhibit and label a few years ago, that the Mackenzie instrument may well be the oldest recording polygraph still in existence. We certainly do not know of any older instrument now in a museum. The description below and the photograph is by W.D. Hackmann.

* * * * * *

349

HISTORICAL NOTE



MACKENZIE'S CLINICAL POLYGRAPH, C. 1910

Unsigned, nickel-plated, dimensions of base 3.7 x 5 cm, and overall height 7.5 cm, in leather box with gold tooling, and marked "Royal Society".

The accurate clockwork of this ink polygraph drives the paper ribbon and (one-fifth second) time marker; it is encased in the nickel-plated housing, from which projects an arm for two registering tambours (one missing), connected by means of rubber tubing to the receiving apparatus - simple aluminum cups for the heart-beat (cardiogram) and venous pulse, and the pelotte, fixed to the arms by straps, for the arterial pulse.

James Mackenzie, the English clinician and cardiologist, described his 'clinical polygraph' in 1892: "inasmuch as the whole arrangement can be used for taking, at the same time and on the same recording surface, tracings of the radial pulse, with tracings of the apex beat, carotid, venous or liver pulse, or the respiratory movements, and as it size is such as to permit its being carried about with the greatest facility, and readily employed in general practice ... ". The first instrument was made for him by Krohne and Seseman of London, but when it became popular, it was manufactured by many firms, including Down Bros. of London and the Cambridge Scientific Instrument Company. Mackenzie was particularly interested in preventive medicine, and concerned that advances in medical instrumentation might not improve medical practice: "While it may be claimed that we may have one hundred new methods for investigating disease in the living ... we have one hundred more ways of going astray."

HISTORICAL NOTE

References Cited

Gay, William O. (1948). The search for truth. <u>English Police</u> <u>Journal</u>, <u>21</u>(4), 284.

Hackmann, W.D. Letter of 16 January 1992.

Macdonald, Arthur (1902). Hearing on the bill (H.R. 14798) to establish a laboratory for the study of the criminal, pauper, and defective classes ... before the Committee on the Judiciary. Washington: Government Printing Office, pp. 176-219. See also Macdonald, Arthur (1905). "Psycho-physical and Anthropological Instruments of Precision in the Laboratory of the Bureau of Education," in Man and Abnormal Man, Including a Study of Children, in Connection with Bills to Establish Laboratories Under Federal and State Governments for the Study of the Criminal, Pauper, and Defective Classes, with Bibliographies, Senate, 58th Congress, 3rd Session, Document No. 187. Washington: Government Printing Office, pp. 163-207.

Mackenzie, James (1908). The ink polygraph. <u>British Medical</u> Journal, <u>1</u>, 1411.

Reid, John E. and Inbau, Fred E. (1977). <u>Truth and Deception</u>: <u>The Polygraph ("Lie Detector")</u> <u>Technique</u>, second edition. Baltimore: Williams & Wilkins, p. 4.

Wygant, James R. Letter of 26 September 1991.

* * * * * *